

## The future of the electronic scientific literature

**The Internet's transformation of scientific communication has only begun, but already much of its promise is within reach. The vision below may change in its detail, but experimentation and lack of dogmatism are undoubtedly the way forward.**

**T**he Internet is easier to invent than to predict" is a maxim that time has proven to be a truism. Much the same might be said of scientific publishing on the Internet, the history of which is littered with failed predictions. Technological advance itself will, of course, bring dramatic changes — and it is a safe bet that bright software minds will punctually overturn any vision. But it is becoming clear that developing common standards will be critical in determining both the speed and extent of progress towards a scientific web.

'Standards' for managing electronic content are hardly a riveting topic for researchers. But they are key to a host of issues that affect scientists, such as searching, data mining, functionality and the creation of stable, long-term archives of research results. Moreover, just as the Internet and web owe their success to agreed network protocols on which others were able to build, common standards in science will provide a foundation for a diversity of publishing models and experiments and be a better alternative to 'one-size-fits-all' solutions.

This explains why the Open Archives Initiative (OAI), one of many alternatives now being offered to scientists to disseminate their work, has now broadened its focus from e-prints to promoting common web standards for digital content.

The reason is that some of the most promising emerging technologies will only realize their full promise if they are adopted in a consensual fashion by entire communities. At the level of the online scientific 'paper', one major change, for example, is a shift in format to make papers more computer-readable. Searches will become much more powerful; tables and figures will cease to be flat, lifeless objects, and instead will be able to be queried and manipulated by users, using suites of online visualization and data-analysis tools.

This is being made possible by Extensible Mark-up Language (XML), which allows a document to be tagged with machine-readable 'metadata', in effect converting it into a sort of mini-database. Most web pages today are coded in HTML. But this contains information only about a page's appearance. Whereas HTML specifies title and author information, for example as simple headings, such as:

```
<H1> The future of the electronic scientific literature </H1>
<H3>by John Smith</H3>
```

XML specifies these in a way that computers can understand:

```
<articletitle> The future of the electronic scientific literature </articletitle>
<author><firstname>John</firstname> <lastname>Smith</lastname>.
```

The possibilities for tagging are endless. But a major need now is for stakeholders to agree on common metadata standards for the basic structure of scientific papers. This would allow more specific queries to be made across large swathes of the literature. Indeed, what is above all hampering the usefulness of today's online journals, e-print archives and scientific digital libraries is the lack of means to federate these resources through unified interfaces.

The OAI has agreed metadata standards to facilitate improved searching across participating archives, which can therefore be queried by users as if they were one seamless site. The OAI is attractive compared with centralized archives in that it allows any group to

create an archive while, by agreeing common standards, they become part of a greater whole. The idea is catching on: it is supported by the Digital Library Federation (DLF), a consortium of US libraries and agencies, including the Online Computer Library Center. CrossRef, a collaboration of 78 learned society and commercial publishers, in which *Nature's* publishers are taking a leading role, is also actively developing common metadata standards that would allow better cross-searching of the 3 million articles they hold.

### Minimal options

As metadata are expensive to create — it is estimated that tagging papers with even minimal metadata can add as much as 40% to costs — OAI is developing its core metadata as a lowest common denominator to avoid putting an excessive burden on those who wish to take part. But even these skimpy metadata already allow one to improve retrieval. This strategy is sensible as it acknowledges the fact that the value and nature of scientific information are heterogeneous.

Minimal metadata will suffice for much of the literature. But there will increasingly be sophisticated and novel forms of publications built around highly organized communities working off large, shared data sets. These hubs will stand out by their large investment in rich metadata and sophisticated databases. The future electronic landscape should see such high added-value hubs evolving as overlays to vast but largely automated literature archives and databases.

In such an early stage of development, it is essential to avoid dogmatic solutions. Not all papers will warrant the costs of marking up with metadata, nor will much of the grey literature, such as conference proceedings or the large internal documentation of government agencies. Many high-cost, low-circulation print journals could be replaced by digital libraries. Overheads would be kept low, and the economics argues that the cheapest means of handling the bulk of the literature may be automated digital libraries. Tags automatically generated from machine analysis of the text, for example, might minimize the quantity of manual metadata needed.

Or take ResearchIndex, software produced by the computer company NEC, which builds digital libraries with little human intervention. It gathers scientific papers from around the web and, using simple rules based on document formatting, can extract the title, abstract, author and references. It interprets the latter, and can conduct automatic citation analyses for all the papers indexed. Such digital libraries will also provide new tools, for example to generate new metrics based on user behaviour, which will complement and even surpass citation rankings and impact factors.

At the other end of the spectrum, specialized communities organized around shared data sets will produce highly sophisticated electronic 'publications', making it much more arduous for authors to submit information because of the amount and detail they will be required to enter in machine-readable form. Take the Alliance for Cellular Signaling (AfCS), a 10-year, multimillion-dollar, multi-disciplinary project run by a consortium of 20 US institutions. It is



taking a systems view of proteins involved in signalling, and integrating large amounts of data into models that will piece together how cellular signalling functions as a whole in the cell. Here, authors would be required to input information, for example, on the protocols, tissues, cell types, specific concentration factors used and the experimental outcomes. Inputs would be chosen from menus of strictly defined terms and ranges, corresponding to predefined knowledge representations and vocabularies for cell signalling.

The idea is that, rather than simply producing their own data, communities instead create a vast, shared pool of well-structured information, and benefit by being able to make much more powerful queries, simulations and data mining. A series of 'molecule pages' would also pull together virtually all published data and literature about individual molecules in relation to signalling.

Indeed, the high-throughput nature of much of modern research means that, increasingly, important results can be fully expressed only in electronic rather than print format. Systems biology in particular is driving research that seeks to describe the function of whole pathways and networks of genes and proteins, and to cover scales ranging from atoms and molecules to organisms. Increasingly, the literature and biological databases will converge to create new forms of publications. Other disciplines stand to benefit, too.

#### Helping machines make sense of science on the web

Many communities, including the AfCS, are building ontologies to underpin such schemes. Ontologies mean different things to different people, but they are in effect representations that attempt to hard-code human knowledge about a topic and the intrinsic relationships in ways that computers can use. The microarray community has been very active in this area. The Microarray Gene Expression Database group has coordinated global standards; as a result, users will be able to query vast shared data sets to find all experiments that use a specified type of biological material, test the effects of a specified treatment or measure the expression of a specified gene, and much more.

One major problem is that genes and proteins often have different names in different organisms, and these often say little about what they do. To get round this problem, the Gene Ontology (GO) Consortium is creating tree-like ontologies of the 'molecular function', 'biological process' and 'cellular component' of gene products. All genes involved in 'DNA repair', for example, would be mapped to the corresponding GO term, irrespective of their name or source organism. A microarray gene-expression analysis that previously yielded only names of expressed genes would in addition carry mapped GO terms that might reveal, say, that half the genes are involved in 'protein folding'. GO terms can also help to federate disparate databases.

Ontologies can also be used to tag literature automatically, and will be particularly useful for grey literature and archival material for which manual tagging was not justified. Papers tagged automatically with concepts can be matched, grouped into topic maps and mined. By breaking down terminological barriers between disciplines, this should also enhance interdisciplinary understanding and even serendipity. *Nature* is actively investigating such possibilities.

The GO ontologies are still very incomplete, however, and the internal relationships need to be enriched. Moreover, caution is required against prematurely pigeon-holing gene functions, given the uncertainty of most annotations. Ontologies are also the focus of intensive research in computing science, and biology is not yet up to speed on this. Efforts such as GO and the Bio-Ontologies Consortium deserve support. Indeed, given the shortcomings of existing ontologies and controlled vocabularies, there may be a case for creating a more organized international effort to ensure economy of effort, interoperability and sharing of expertise.

The advent of structured papers that are increasingly held in literature databases blurs further the distinction between the scientific paper and entries in biological databases. Already, entries in the biological databases are often hyperlinked to relevant articles in the literature and vice versa, and CrossRef is developing standards for such link-

ing. As text becomes more structured, it will be possible to increase the sophistication of both linking, data manipulation and retrieval.

Biological databases and journals have evolved relatively independently of one another. Database annotations lack the prestige of published papers; indeed, their value is largely ignored by citation metrics, and their upkeep is often regarded as a thankless task. Database curation has consequently lacked the quality control typical of good journals. The convergence between databases and the literature means that database annotators and curators will increasingly perform the functions of journal editors and reviewers, while publishers will develop sophisticated database platforms and tools.

#### New ways in

Database- and metadata-driven systems will drive interfaces to publications from simple keyword search models to ones that reflect the structure of biological information. Visualization tools of chromosomal location, biochemical pathways and structural interactions may become the obvious portals to the wider literature, given that there are far fewer protein structures or gene sequences than there are articles about them. As Mark Gerstein, a bioinformaticist at Yale University, points out: "One might 'fly through' a large three-dimensional molecular structure, such as the ribosome, where various surface patches would be linked to publications describing associated chemical binding studies."

Future electronic literature will therefore be much more heterogeneous than the current journal system, and dogmatic solutions should therefore be resisted. It is significant and sensible that both CrossRef and OAI have made key strategic choices favouring openness and adaptability. They seek to federate distributed actors rather than to create centralized structures. They also make their work independent of the type of content, which makes it flexible enough to incorporate and link seamlessly not just papers but news, books and other media.

Crucially, both OAI and CrossRef have also decided to build systems independent of the economic mechanisms surrounding that content. Many publishers, in particular some learned societies, may be willing to make their content free, perhaps after a certain delay. Others are exploring business models where authors or sponsors pay, which would allow free access to articles on publication. The open technological frameworks also mean that particular communities, such as scientists with specific metadata needs for their discipline, are free to build in more complex data structures; the higher overheads incurred may require charging for added-value services.

#### Neutrality

The OAI and CrossRef strategies therefore differ fundamentally from more centralized systems proposed by PubMed Central (PMC), operated by the US National Library of Medicine, and E-Biosci, being developed by the European Molecular Biology Organization.

But PMC and E-Biosci highlight the urgent need to index the full text of papers and their metadata and not just abstracts, as is the practice of PubMed and other aggregators. Services that require publishers to deposit full text only for indexing and improving search are useful.

Unfortunately, PMC, unlike E-Biosci, confounds this primarily technological issue with an economic one, by requiring that all text be made available free after, at most, one year. It is regrettable that PMC has not in the first instance sought full-text indexing itself as a goal, as this in itself would be an immediate boon to researchers. It would also probably have been more successful in attracting publishers.

The reality is that all of those involved in scientific publishing are in a period of intense experimentation, the outcome of which is difficult to predict. Getting there will require novel forms of collaboration between publishers, databases, digital libraries and other stakeholders. It would be unwise to put all of one's eggs in the basket of any one economic or technological 'solution'. Diversity is the best bet. ■

The web version of this Opinion article contains hypertext links to the named organizations and to related articles in *Nature's* web forum on access to the electronic literature at <http://www.nature.com/nature/debates/e-access/>



**Male forensics**  
Y's and wherefores  
of investigating sex  
crimes  
p6



**Law force**  
US military turns  
its guns on 'green'  
laws  
p7



**A different class**  
Japan puts personal  
emphasis on science  
teaching  
p8



**Viral vision**  
Concerns arise  
over gene-therapy  
technique  
p9

# Fears for basic science as Bush backs use of investment criteria

**Colin Macilwain, Washington**

The White House has fuelled growing fears among scientific leaders that the recent era of expansion in US science funding is coming to a close. On 25 August, it announced plans to set new "investment criteria" next spring for all research programmes, including those in basic science.

Reports that the National Science Foundation (NSF) and the Office of Science at the Department of Energy (DOE) are planning to cut their budgets for 2003 have added to scientists' worries.

Eight months into the Bush administration, scientific leaders are also increasingly concerned about their lack of access to the White House. John Marburger, director of the Brookhaven National Laboratory in New York state, who was named as Bush's science adviser in June, has not come to Washington as a consultant, as the science community would have liked, in advance of his expected confirmation by the Senate next month.

No other senior research position in the administration has yet been filled, and the three most senior officials left from the Clinton administration — NSF director Rita Colwell, NASA administrator Dan Goldin, and Richard Klausner, director of the National Cancer Institute — are each tipped to leave their positions soon. So is energy secretary



Spencer Abraham — sworn in this January — who may run for governor of Michigan.

Bush's plan, released by the White House Office of Management and Budget (OMB), says the criteria-based approach is being tested now on the DOE's applied research programmes in energy supply and conservation.

The OMB will next confer with the five main science agencies — the National Institutes of Health, the NSF, the DOE, NASA and the Department of Defense — before extending the approach to all basic research programmes. The chosen criteria will be



**Mitch Daniels (above) and George Bush plan new performance measures for basic research.**

published in the spring and will be used to develop budgets for the 2004 fiscal year.

Science lobbyists fear that the proposal will hurt basic science by demanding quick pay-offs. One senior lobbyist says of the officials driving the new policy, who include Mitch Daniels, the OMB's director: "They don't know what basic research is. They think that it should all be done by industry."

But White House officials argue that the OMB will listen to scientists' views before establishing the criteria for basic research. ■

♦ <http://www.whitehouse.gov/omb/budget/fy2002>

## Stanford gift scaled back over federal stem-cell policy

**Rex Dalton, San Diego**

A technology entrepreneur has withdrawn \$60 million of a \$150-million gift for a biomedical research facility at Stanford University in California, in protest at US policy on stem-cell research.

James Clark — a former Stanford professor who made his fortune founding Silicon Graphics and Netscape — made his announcement in *The New York Times* on 31 August. He said he would hold the gift to \$90 million because of the Bush administration's decision to limit stem-cell research funding to a small number of cell lines already in

existence (see *Nature* 412, 665; 2001).

Clark wrote that this plan was "beyond comprehension" and "driven by ignorance, conservative thinking and fear of the unknown".

Clark had pledged to build a new centre for biomedical engineering and science — known as Bio-X — at Stanford. The centre was conceived to encourage collaboration between Stanford's schools for biology, engineering and computer science.

Construction of Bio-X began in the summer, and Stanford officials say they can still complete the building, but that Clark's

action will delay funding for research projects at the centre. It is understood that Clark had originally planned to withdraw even more of his gift, halting construction entirely, but was dissuaded from doing so.

The withholding is the latest difficulty faced during the three years of planning for Bio-X. The building's size was scaled back to appease the surrounding community of Palo Alto. And in June, project director James Spudich said he would step aside this autumn. But after Clark's announcement, Spudich said he will remain as director for as long as is needed. ■



# Y-chromosome analysis urged for sex crimes

Quirin Schiermeier, Munich

Forensic scientists say that methods that make use of the Y chromosome will soon provide a valuable new generation of tools for investigating sex crimes and settling paternity suits.

Its male specificity and pronounced variation make the Y chromosome a useful tool for forensic geneticists. This is particularly true in difficult cases of sexual assault where the DNA samples from offenders and victims may have been mixed in a way that makes them difficult to separate for conventional DNA analysis.

At the biannual meeting of the International Society for Forensic Genetics, held in Münster, Germany, late last month, scientists announced plans to enlarge and improve existing reference databases of Y-chromosome distributions in European and US male populations. Data from Asia and South America are also to be added to the database, they said.

Unlike the gender-independent chromosomes normally used by forensic geneticists, the Y chromosome carries mutations over many generations, providing unambiguous male lineages. To differentiate chromosomes, geneticists use haplotypes, which are chunks of DNA containing closely linked genetic variations that were inherited as a unit.

A standardized European reference database was created in 1995 by geneticists at the Institute of Legal Medicine of the Humboldt University in Berlin. It is based on a set of nine

highly variable markers on the Y chromosome, and allows forensic researchers to extrapolate the frequency of a given haplotype in a sufficiently homogeneous population, such as that in much of Europe.

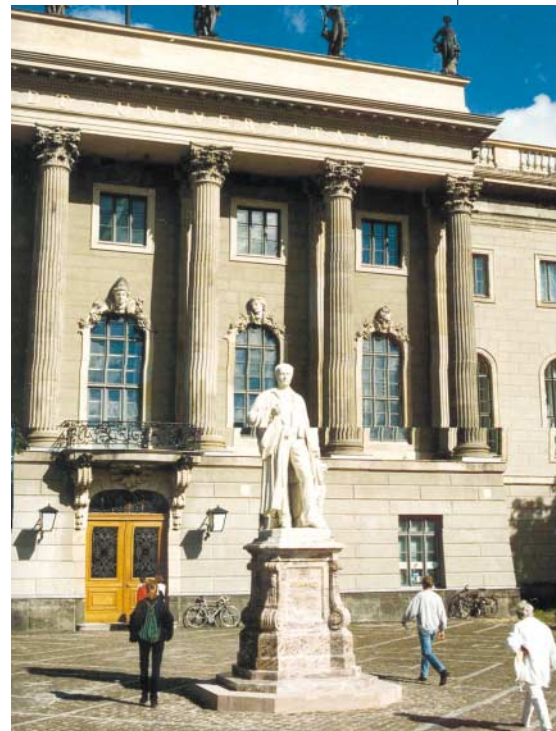
The database currently contains around 7,000 haplotypes from 51 European population samples, gathered from 40 forensic institutes across the continent.

A Y-chromosome haplotype can help to identify an offender or determine fatherhood with a probability of 99.9% or more, says Lutz Roewer, one of the database's founders. The results can be used in court as additional evidence, although they are less conclusive than conventional DNA analysis of markers on non-sex-specific chromosomes.

Y-chromosome analysis could also help to determine, or exclude, the ethnological background of male suspects — raising serious ethical and political questions about its use. As Roewer points out: "Even if forensic geneticists were at a stage to tell the police not to search for a rapist among, for example, the white European population, it would be ethically problematic."

Apart from forensic applications, scientists believe that a well-maintained and continuously growing Y-chromosome database could be valuable in many areas, such as population history, migration research and evolutionary genetics.

"The Y chromosome is extremely interesting for us because of its variability and



Humboldt University's Institute of Legal Medicine, the origin of the European Y database.

high frequency of tolerated mutations," says Werner Schempp, an anthropologist and human geneticist at the Albert Ludwigs University in Freiburg, Germany. ■

♦ [www.ystr.org/europe](http://www.ystr.org/europe)

♦ [www.ystr.org/usa](http://www.ystr.org/usa)

CHARITÉ, HUMBOLDT UNIVERSITY

# Journal boycott presses demand for free access

Jonathan Knight, San Francisco

Researchers began a boycott of scientific journals that do not allow free access to their contents on 1 September — although organizers admit that too few journals have complied for the boycott to take full effect immediately.

The organizers, members of an initiative known as the Public Library of Science (see *Nature* 410, 502; 2001), say the list of compliant journals "is not yet sufficient to accommodate all the work that we and our colleagues must publish".

In a letter to the 26,000 researchers who have pledged support on the Internet, they urged supporters not to subscribe to, publish in or review for journals — including *Nature* — that do not make their contents freely available within six months of publication on a centralized website, PubMed Central, operated by the US government. But they suggest that, if necessary, supporters should publish

in journals that come closest to meeting its conditions.

At the time of the boycott deadline, 16 print journals and 60 online journals published by London-based BioMed Central met its criteria.

Print journals complying with the criteria include *Proceedings of the National Academy of Sciences (PNAS)*, *Molecular Biology of the Cell*, *British Medical Journal*, *Bioinformatics*, *The Plant Cell* and *Breast Cancer Research*.

The letter confirms earlier reports that the organizers are trying to raise funds to start their own online journals (see *Nature* 412, 469; 2001). The journals would be produced by volunteers, and the group says their publication cost could be covered by a flat fee of about US\$300 per published article.

Whether such journals will draw researchers away from mainstream journals is hard to predict, says Donald Kennedy, editor of *Science*, which makes its

contents freely available on its own website after a year.

"A lot will depend on whether a large enough group of people will support it and provide all of the other things *Nature* and *Science* provide, such as commentary and news," he says.

Nicholas Cozzarelli, editor-in-chief of *PNAS*, says he strongly supports the group. "I have told them I'll be glad to pitch in, and I think you'll find they will get a tremendous number of volunteers," he says.

Jayne Marks, publishing director of the Nature Publishing Group, which does not make its full content available free, expressed confidence that its journals, which include *Nature*, would continue to receive high-quality submissions. "People judge the publication on what it publishes and the benefits it provides," she says. "We are doing a lot to enhance access to the literature through collaboration with other publishers and archives." ■

# Environmental laws face military manoeuvres

**Mark Schroppe**

The US military is mounting an offensive this year against an enemy within: the environmental regulation that, the Pentagon argues, constrains its training and affects military readiness.

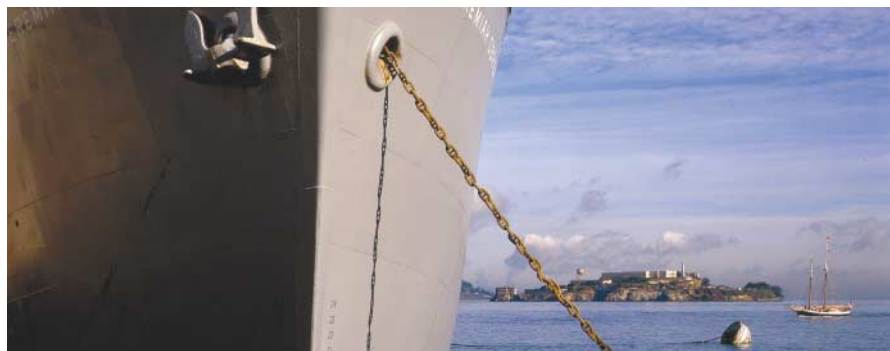
The armed forces and their allies in Congress are all pushing for more exemption from environmental laws, arguing that some of them damage national security.

Ecologists and marine biologists, together with environmental groups, regard the laws as important tools in conserving the ecosystems on large tracts of land owned by the military, and in protecting marine wildlife from damage caused by war games. And although the Pentagon is yet to endorse any legal changes, it may be looking to the conservative administration of George W. Bush to create exemptions from laws such as the Endangered Species Act and the Marine Mammal Protection Act.

Last month, in a bid to expose this intention, an environmental group leaked a draft of a Navy document outlining problems that stem from environmental laws. The document, which has been in development since last December, describes problems such as the Navy's continuing struggle to test a sonar that might harm marine mammals (see *Nature* **410**, 505; 2001), and lists ambiguities in current laws and their enforcement. It calls for more clarity in environmental laws, and for the collation and publication of data on their financial and operational impacts.

"This is what we would expect to see from the Bush administration in its attempts to roll back the environmental gains of the past 10 years," says Dan Meyer, a former Navy lieutenant and now a lawyer at Public Employees for Environmental Responsibility, the Washington-based lobby group that released the document.

Doug Spencer, a spokesman for the Navy, says: "The military is not out to damage the environment." He adds that it is against Navy policy to discuss working drafts, but that a



**Legal anchor:** a leaked Navy document outlined problems linked to US environmental regulations.

final version of the paper is now complete and will shortly be sent to Congress.

At a hearing of a House Armed Services subcommittee earlier this year, Joseph Angelo, a senior Pentagon official, said that unrealistic combat training had "cost us lives and equipment" during real wars, such as the recent conflict in Kosovo. He argued that environmental rules restrict military exercises, causing "a slow degradation in our ability to test and train effectively". He also noted the irony that restrictions on base activities have increased as a result of the military's success in managing these areas — to the extent that some bases have become critical havens for endangered species.

The House of Representatives is consid-

ering a law that asks the Department of Defense to conduct national-security assessments on proposed operations, to accompany the environmental-impact assessments that are already required. Spencer says this would give a more balanced view of potential conflicts. "We have to look at the risks to national security," he says.

James MacMahon, an ecologist at Utah State University and former president of the Ecological Society of America, says he can imagine few situations in which the military would be unable to achieve a given goal without affecting endangered species. Any exemptions should be based on a national-security assessment of the potential number of lives the activity could save, he suggests. ■

## Listing resumes for species at risk

**Mark Schroppe**

The listing of endangered species of plants and animals in the United States is to be revived, following an agreement between the government and environmental groups.

For the past year, most of the government's budget for listing newly endangered species has been used to fight litigation from environmental groups aimed at forcing it to designate habitat needed for the recovery of species already listed.

In October 2000 the US Fish and Wildlife Service said that such legal work would probably consume its entire \$6-million listing budget for 2001, leaving nothing for the field assessment and listing of new species. Since then, only a few new species have been listed, mostly in response to court orders or to prevent litigation.

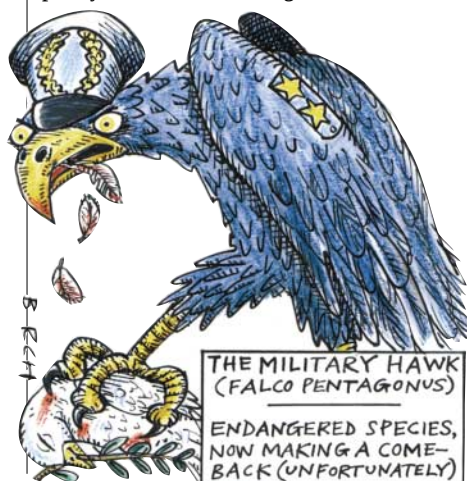
Faced with this, several environmental groups that were suing the service have been negotiating with it instead. Under an agreement announced on 29 August, the groups will allow the government more time to comply with court orders requiring it to designate habitat for eight protected species.

The government will in turn expedite the listing process for 29 of the more than 200 species under consideration.

The 29 include three species whose protection is considered an emergency, such as the pygmy rabbit, *Brachylagus idahoensis*, in Washington state, which has a population of less than 50. "It's a common-sense solution to the bad situation we found ourselves in this year," says Gary Frazer, assistant director for endangered species with the service.

Species were chosen according to such factors as their degree of risk and whether necessary protection measures are known, according to Kieran Suckling, head of the Center for Biological Diversity, based in Tucson, Arizona, one of the groups involved.

"We're biting off the tiniest tip of the iceberg," says Suckling. "To address the problem adequately, the Fish and Wildlife Service has to have a budget that is up to the task." The proposed budget for 2002 includes \$8.5 million for listing species, although the service's officials say it needs \$120 million to meet all of its legal liabilities under the endangered species act. ■





# Japan plans revision of science lessons

David Cyranoski, Tokyo

Alarmed by declining student interest in science and the superior test scores of some of its neighbours in east Asia, Japan is planning to reform the way science is taught in its schools.

"In the past we focused on making sure that everyone was getting a standard education," says a government official involved in the reforms. "Now we want to focus on enhancing the respective talents of each student."

The plan will start next year with the establishment of 20 'super-science high schools', each of which will receive special funding to buy equipment and hire teaching assistance from university researchers. The schools will also form 'science clubs' and confer with the universities on teaching methods and curriculum content. In addition, the plan will provide small grants for equipment to 1,500 other schools.

Yoza Shimomura, an Earth sciences teacher at Shinagawa Etoile Girls' High School in Tokyo, says he hopes that the plan will help to encourage independent thought among students. But critics have expressed concern that overworked teachers in Japan's



Learning shortfall: children's test scores in Singapore, Taiwan and Korea are surpassing those in Japan.

large classrooms need far more help than the package of measures is likely to offer.

Standardized tests of 13–14-year-olds indicate that children in Singapore, Taiwan and Korea are doing better than those in Japan. And surveys conducted at the same time as the tests found that the number of

Japanese students who enjoy science or would like to make a career out of it is among the lowest in the countries tested.

Japan's newly formed Ministry of Education, Culture, Sports, Science and Technology is seeking ¥9 billion (US\$75 million) in its budget request for the reforms. ■

## Ministry attempts to breathe life into clinical trials

David Cyranoski, Tokyo

The Japanese government has unveiled fresh incentives that it hopes will encourage doctors and researchers to participate in clinical trials of combination therapies.

According to statistics from the Ministry of Health, Labour and Welfare, the number

of trials sponsored by the drugs industry in Japan has fallen by half over the past five years. The ministry is to introduce a grant scheme that will provide ¥10 billion (US\$84 million) next year for clinical trials.

Toshiro Nakagaki, an official in the health ministry, says that the new scheme will help to reverse this decline and will build up Japan's capacity for conducting drug trials. The grants, which will probably last for three years, will be available to teams of doctors working on various diseases, including cancer, heart disease and diabetes.

The fall in the number of clinical trials is blamed on several botched investigations. For example, in 1993, three patients died during a trial for the shingles treatment sorivudine — and a further 15 died after the drug was approved — because it reacted with their fluorouracil anticancer treatment.

Tough new procedures for informed consent, implemented in 1998, have also hampered the organization of clinical trials. Doctors who take part have to work through the necessary paperwork during their already-crowded working hours, says Shigenobu Kanba, a neuropsychiatrist at Yamanashi Medical University.

And the trials can take up to five or six years to complete, according to Kanba. As a

result, many drug developers turn to contract research organizations that can do tests more quickly and cheaply abroad. "Some companies do all their research overseas," says Keiko Oishi, head of international relations at CMIC, a Tokyo-based company that contracts out clinical trials on behalf of pharmaceutical companies.

The past couple of years have shown some signs of improvement. New patient-recruitment centres and the recent legalization of recruitment advertisements have produced more willing patients for trials. "Increasingly, companies are realizing that for practical reasons, it is necessary and possible to do trials in Japan," says Oishi.

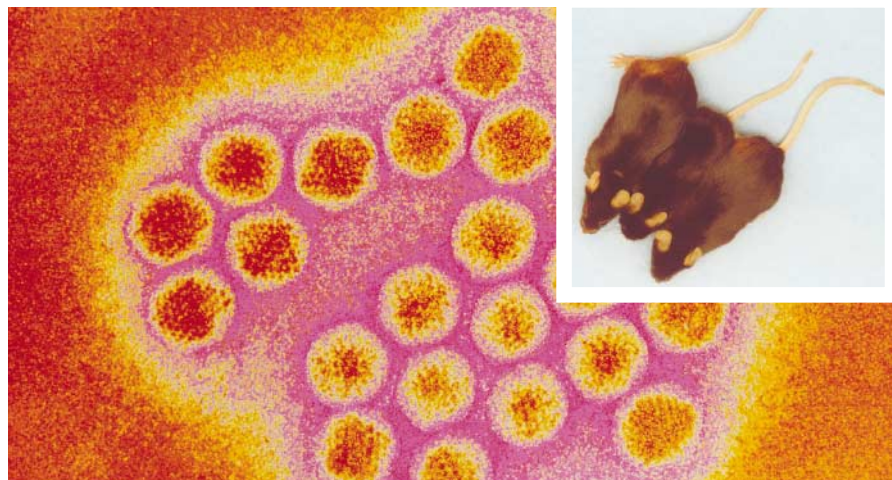
But the key problem is doctors' reluctance to participate. "Until now, doctors have thought of clinical trials as a lowly part-time job, and papers or successful trials don't add to the list of a doctor's achievements," says Kanba.

The grants to be made available under the new ministry programme could help, says Oishi. They can be used, for example, to pay for clinical-research coordinators who will help to recruit patients and take care of informed-consent procedures. They will also finance any extra tests the doctors might have to perform during the trials. ■



The scheme could encourage trials at institutes such as Osaka University's medical school.

# Liver tumours temper hopes for gene-therapy technique



Is it safe for humans to receive genes from adenoviruses like those used to treat mice?

## Helen Pearson

A promising approach to gene therapy received an important fillip this week with the publication of two papers suggesting that it can effectively treat rare diseases in mice. But a third paper raises questions about the safety of the technique after tumours were found in older mice that had been used in one of the studies.

The existence of the tumours, whose cause remains unclear, has led to calls for more long-term toxicity studies in animals to be completed before such techniques are used in human trials.

Their discovery, by researchers at Washington University in St Louis, Missouri, is reported in this month's edition of *Gene Therapy* alongside two sets of positive findings on the use of the technique — one from the Washington University group, the other from researchers at the University of Florida at Gainesville. The two groups joined forces to publish the cautionary paper.

The technique in question uses an adeno-associated virus (AAV), which is not linked to human disease and is therefore considered relatively safe, to deliver normal genes that can substitute for genes containing disease-causing mutations. It is regarded as one of the most exciting prospective approaches to gene therapy.

Last December, a team led by Mark Sands of Washington University found liver tumours in 6 out of 59 mice, 18 months after they had received a single AAV injection for the rare genetic disease mucopolysaccharidosis type VII. At the time, says Sands, "it got me really nervous".

Despite the discovery of the liver tumours, the Food and Drug Administration (FDA) has given permission for the use of AAV to continue in human clinical

trials for cystic fibrosis and haemophilia B.

The US authorities have been cautious about gene-therapy experiments since the death of Jesse Gelsinger in a gene-therapy trial at the University of Pennsylvania in September 1999. Sands' findings initially prompted an immediate halt of human clinical trials using AAV last December. The trials were reinstated by the FDA in January, on the grounds that other research groups working with AAV had failed to find similar tumours in their older animals.

Some think the authorities overreacted in halting the trials in the first place. "It was a bit drastic," says AAV researcher Robin Ali of the Institute of Ophthalmology in London.

Others say the FDA was right to exercise diligence. "You have to take any theoretical risk extremely seriously," says Duncan Geddes, a cystic-fibrosis clinician at the Royal Brompton Hospital in London.

The assessment of the tumour problem published by Sands' group suggests that it was unlikely to have been caused by AAV, because the researchers had found insufficient sign of the virus DNA. But the authors say that they cannot entirely exclude the possibility that AAV caused the tumours.

The latest result is unlikely to be the last word on discussions about the safety of AAV. Researchers say that more priority should be given to long-term toxicity studies of such treatments in animals. These expensive studies require thousands of animals, and often take place only after initial human trials are already under way.

"It might be worthwhile doing them earlier," suggests Terry Flotte of the University of Florida, one of the authors of the second study that found the technique to be effective in mice.

► <http://www.naturesj.com/gt>

# Celera and Motorola brought in to aid hunt for disease genes

## Jonathan Knight, San Francisco

A public-private research consortium has contracted two companies, Motorola and Celera, to develop a tool to find genes involved in heritable diseases.

The tool is a statistical map of inheritance patterns of single-nucleotide polymorphisms (SNPs), the sequence variants scattered throughout the genome that account for most genetic differences between individuals. The information will help researchers find areas of the genome associated with hereditary diseases.

The SNP Consortium — a collaboration of university sequencing centres, ten pharmaceutical companies and the Wellcome Trust, a British medical research charity — has so far compiled a public list of almost 2 million SNPs.

Because they vary from person to person, SNPs are effective markers for tracking genetic regions through family histories. But because chromosomes exchange DNA with each generation, researchers need to know the probability of nearby SNPs staying together.

In the collaboration, announced on 28 August, the sequencing firm Celera, of Rockville, Maryland, and the life sciences unit of Motorola, in Deerfield, Illinois, will chart inheritance patterns of 2,000 SNPs in 650 people from 50 families. Although Celera claims to have identified more than 3 million SNPs, only public data will be used in this project. Financial details have not been disclosed.

The map is due to be completed by the end of this year. A plan announced in July to produce a much more detailed map of up to 1.5 million SNPs is expected to take several years (see *Nature* 412, 105; 2001).

The small number of SNPs being mapped in the Celera-Motorola project will be useful only for studies within families. Studies of the population at large, in which genetic recombination has been much more extensive, will require larger numbers of SNPs.

Current linkage maps use another genetic variant, microsatellites, which have been very effective, says David Altshuler, a geneticist at Harvard Medical School and leader of a team that published 1.4 million SNPs earlier this year.

He expects people to jump on the SNP bandwagon as technology for detecting them improves. "No one is pouring a lot of money into making microsatellite genotyping cheaper and easier," he says. "But a lot of companies are making SNP genotyping cheaper and easier."



## Feathers fly at claim that Roslin geneticist stole chicken data

**London** The Roslin Institute, the Edinburgh-based genetics centre where Dolly the sheep was created, has been accused of misappropriating research belonging to a US biotechnology company.

AviGenics, a company based in Athens, Georgia, that specializes in avian genetics, has accused Roslin researcher Helen Sang of stealing information about an AviGenics project to create chickens that express human antibodies in the whites of their eggs.

Sang, an expert in gene transfer in chickens, was on the advisory board of AviGenics until late last year. She left shortly before Roslin announced plans for a similar venture with Viragen, a Florida-based biotechnology company. AviGenics has asked the Georgia courts to stop Viragen and Roslin from proceeding with their plans.

AviGenics declined to comment on the case, but Roslin and Viragen both dismissed the action. "Our legal team has reviewed the claims and found that they are totally frivolous," says Doug Calder, director of communications at Viragen.

## Relief as Japan's H-IIA rocket soars at last

**Tokyo** Japan's space scientists breathed a sigh of relief last week when the troubled H-IIA rocket was launched successfully.

The H-IIA's engines failed to fire during tests last summer, delaying plans for a test launch. The rocket is a new version of the H-II, which failed twice in the late 1990s, causing the loss of expensive satellite cargoes.

The H-IIA finally took to the skies on 29 August, carrying an evaluation payload to check the rocket's performance. Japan's National Space Development Agency says it now plans another test flight next year, followed by the launch of an Earth-observing satellite on the rocket in 2003.

♦ [http://www.nasda.go.jp/index\\_e.html](http://www.nasda.go.jp/index_e.html)

## Russian programmer charged in US

**San Francisco** A Russian computer scientist has been charged in California with five counts of copyright violations.

Dmitry Sklyarov, a doctoral student at Moscow State Technical University, and his employer, ElcomSoft, are accused of selling software that allows users to circumvent the encryption technology that protects electronic books. ElcomSoft had made the 'Advanced eBook Processor' software available on its website for \$99.

Sklyarov was arrested in June after he flew to Las Vegas to give a talk on computer-code

## Massive artwork celebrates human genome



The Wellcome Trust has gone to great lengths to brighten up its new headquarters in London. Jason Middlebrook's *The Geology of Biology*, Britain's longest

artwork at 112 metres long, will be unveiled later this month. It was inspired by the publicly funded human genome project, to which the trust was a major contributor.

security at DefCon, a hackers' conference.

The arrest has prompted protest by computer scientists and hackers in Moscow and other cities around the world.

"He is a scientist, a programmer," says Sklyarov's lawyer Joseph Burton. "The critical question is: will scientists around the United States and the world face legal prosecution for doing something they consider legal? That is pretty scary."

## Heart-transplant pioneer Barnard dies at 78

**London** Christiaan Barnard, the first surgeon to perform a successful human heart transplant, died last week, aged 78. He became internationally famous after the operation at a hospital in Cape Town, South Africa, in 1967.

In the following years, Barnard was forced to defend heart transplants against people who believed the heart to be the seat of the emotions, and against a more general concern that swapping body parts interfered with a divine plan. The heart's job is to pump blood, not emotions, he insisted.

Barnard's own heart and emotions left a trail of glamorous, if broken, relationships. But he was highly respected for both his anti-apartheid stance and his support for work with endangered species.

In later years he continued to be controversial, claiming, for example, that human cloning could be used to avoid the defects produced by nature.

## \$20 million to find new ways of teaching

**Washington** University researchers are being offered the chance to apply for US\$1 million grants from the Howard Hughes Medical Institute to help them develop innovative approaches to undergraduate teaching.

"Research is advancing at a breakneck pace, but many college students are still listening to lectures in large classes and memorizing facts from textbooks," says Tom Cech, the institute's president. The institute hopes to create a select group of scientist/

educators who will become leaders in their fields, but is keeping an open mind about the schemes to support. Twenty grants of \$250,000 a year for four years are available. ♦ <http://www.hhmi.org>

## Medical centre launches lawsuit against Genentech

**Los Angeles** The San Francisco-based biotechnology firm Genentech has been taken to court, accused of underpaying a non-profit medical centre from which it licensed rights to produce an insulin drug.

Scientists at the City of Hope National Medical Center, outside Los Angeles, discovered how to make recombinant human insulin some 23 years ago. Genentech licensed the rights to the discovery and marketed an insulin drug based on it. The centre now claims that Genentech has been hiding revenue it earned from the drug.

"The City of Hope and Genentech have had a good working relationship for more than 20 years, during which the City of Hope accepted \$285 million in royalty payments," says Genentech spokeswoman Sabrina Johnson. "Now they are trying to rewrite a more-than-20-year-old contract."

## Toxic cowpats poison Alpine meadows

**Paris** French scientists met in Isère in south-eastern France last week to tackle the problem of toxic cowpats, which are damaging Alpine pasture.

The toxicity is thought to be caused by use of the drug ivermectin, which is administered to cows to protect them from parasites during the summer months, when the animals are in the pasture.

Cowpats produced by the treated cattle are poisonous to dung-eating insects, threatening the insects themselves and the small animals that feed on them. The dung can remain in the pasture for four years or more, preventing plant growth.

The researchers hope to convince the farming community to use alternative treatments in the future.

# Biology's last taboo

Will gene therapy ever extend to inducing changes in humans that can be inherited down through generations? Maybe

so, if the concerns over safety can be ironed out. Jonathan Knight considers the technical challenges and the ethical arguments.



P. PLALLY/EURELIOS/SPL

For a few days in January, a young rhesus monkey called ANDi became one of the most famous animals on the planet, his endearing features plastered across countless newspapers. Engineered by a team at the Oregon Regional Primate Research Center near Portland so that each of his cells carries a gene for a glowing green protein, ANDi is the world's first transgenic primate<sup>1</sup>. His name derives from 'inserted DNA' spelt backwards.

Some reports suggested that ANDi is just one step removed from 'designer' human babies. But if anything, he shows that transgenic people are still far from a practical reality. ANDi gives out not even the faintest green glow — the introduced gene was active in two other fetuses that aborted spontaneously, but in ANDi it seems to lie silent. Nor do ANDi's creators hope to develop a method for use in people. Rather, they are interested in creating genetically modified monkeys for use in medical research. "We discourage any extrapolations to people," says Gerald Schatten, the team's leader, now at the University of Pittsburgh.

But when asked to look a decade or more into the future, many experts believe that 'germline' gene therapy, which aims to produce babies carrying altered genes that will be inherited down the generations, might become part of mainstream medicine. Given advances in basic research, they argue, the safety concerns that currently put germline gene therapy off-limits might be overcome.

Objections to human germline manipulation run deeper than the safety issues. For some people, tampering with our genetic inheritance in this way is fundamentally wrong. Others fear that what starts as an attempt to cure inherited diseases will soon mutate into a programme of genetic

'improvement'. But previous experience suggests that such concerns may not restrain those intent on pushing back the boundaries of reproductive and genetic medicine (see 'The unregulated frontier', overleaf). "Altering the genome is essentially the endpoint of the whole genomic revolution," claims Gregory Stock of the University of California, Los Angeles, an expert on the societal impact of emerging genetic technologies.

## Tackling the taboo

In March 1998, Stock organized a conference in Los Angeles that threw a spotlight on human germline manipulation, then considered a taboo subject. At the meeting, leading scientists spoke in favour of the

idea, arguing that excessive regulation could hamper research of medical value<sup>2</sup>. Last year, a report from the American Association for the Advancement of Science dampened the enthusiasm, calling for a moratorium on attempts to change a person's genes in ways that could affect future generations<sup>3</sup>. But it left the door open: with safe methods, proper supervision and adequate public discussion, the report concluded that germline gene therapy might one day be acceptable.

Any future attempt at human germline gene therapy will draw on experience with transgenic animals. The most basic procedure, injecting DNA fragments into one of the unfused nuclei in a newly fertilized egg, was pioneered in mice more than two decades ago<sup>4</sup>. The



NEWSMAKERS

Monkey business: ANDi, the first genetically modified primate, was born containing a jellyfish gene.





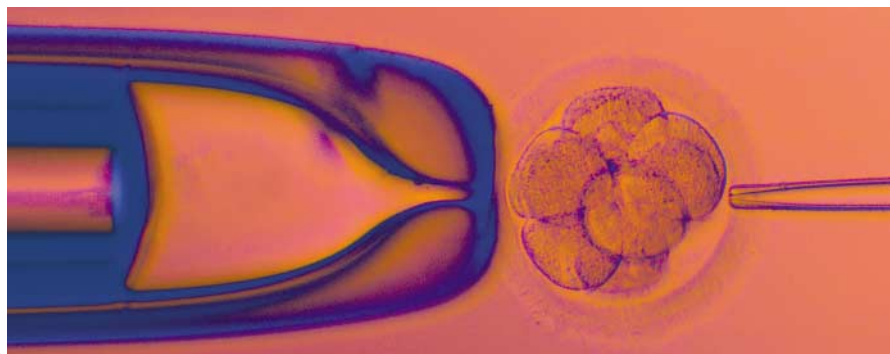
**Gregory Stock sees altering the genome as the 'endpoint' of the genomic revolution.**

inserted DNA only occasionally works its way into a chromosome. Even when it does, it may fail to switch on — of the embryos that survive to be transferred to a surrogate mother, about one in a hundred incorporates a working copy of the gene. At present, there is no way to be sure of the outcome before transferring the embryo to the womb.

In 1998, Robert Bremel and his colleagues at the University of Wisconsin in Madison devised a way to boost the efficiency of transgenesis by using a retrovirus to deliver the genes<sup>5</sup>. Retroviruses insert their genetic material into their host's chromosomes, before hijacking its cellular machinery to make copies of themselves. By modifying a retrovirus so that it could still carry a transgene into the genome of unfertilized cows' eggs, but disabling its ability to replicate, the researchers found that all four embryos that developed to term carried a working copy of the transgene. This technique was used by the team that created ANDi. Researchers in Hawaii have since achieved similar efficiency in mice by injecting eggs with sperm mixed with the transgene<sup>6</sup>.

But there is a bigger problem with techniques that rely on inserting genes into newly fertilized eggs: the consequences are not always easy to predict. When researchers at the US Department of Agriculture added a growth-hormone gene to pigs they got fast-growing livestock. But most of the animals suffered lethargy, arthritis and gastric ulcers — and the boars lost their sex drive<sup>7</sup>.

Such problems are probably caused by the



**Fine tuning: an eight-celled human embryo, created by IVE, has one cell removed for screening.**

random insertion of the transgene into the genome. Multiple copies might be incorporated, and transgenes may also land next to regulatory sequences that suppress or enhance their normal activity or cause them to be expressed in the wrong tissues. Worse, random insertion can also lead to the disruption of other important genes.

### On target

Any attempt to repair genes in human embryos would require a way to ensure that a single copy of the transgene is incorporated in exactly the right position, neatly displacing the target 'faulty' gene. In mice, such gene targeting is now standard practice. It relies on a phenomenon called homologous recombination, in which a gene flanked by sequences that match those bordering a target site in the genome is occasionally inserted in that position by the cell's DNA-repair enzymes, replacing the target gene.

Homologous recombination is a rare event. But in the mid-1980s, the pioneers of gene targeting hit on the idea of adding the DNA to cultured mouse embryonic stem (ES) cells, selecting those in which the gene had incorporated, and then injecting these cells into mouse embryos. Because ES cells can develop into any type of cell in the body, this results in mouse 'chimaeras' consisting

of some cells derived from the original embryo and others from the ES cells. These mice can then be mated with one another to produce pure transgenic mice<sup>8</sup>.

But producing chimaeras is hardly an acceptable scenario for human reproduction. So if society ever decides that tinkering with the human genome is desirable, we will need a new technology. What might it be?

One possibility is an artificial chromosome. Several researchers are manufacturing chromosomes that can function in mouse and human cells. These consist of three basic components: the centromere, the central region to which the microtubules that pull chromosomes apart during cell division attach; the telomeres, which cap chromosome ends; and the intervening DNA that carries genes. Placing transgenes in artificial chromosomes means there is virtually no limit to the size of the genes inserted. It also carries no risk of genetic disruption caused by random insertions into the genome.

The technology, still in its infancy, is being developed for transgenic animal production and gene therapy, rather than for human germline modification. Huntington Willard, director of the Research Institute of University Hospitals in Cleveland, Ohio, and his colleagues at Case Western Reserve University, also in Cleveland, have engineered an artificial human chromosome from



**Handy: a transgenic mouse with no dystrophine in its muscles, created for muscular dystrophy research.**





**Taking a gamble: is the technique used to produce gene-targeted lambs at Roslin the way forward?**

scratch, using only synthetic DNA. When they inject their artificial centromeres together with gene-bearing DNA fragments into human cells, the chromosomes assemble spontaneously and seem to behave like the real thing<sup>9</sup>.

Meanwhile, the Canadian company Chromos in Burnaby, British Columbia, has built artificial chromosomes around portions of mouse centromeres. The chromosomes replicate normally in cells from several species, including cows and humans<sup>10</sup>. Transgenic mice can be created by simply injecting an artificial chromosome into a fertilized egg, and these mice can subsequently pass the chromosome to their offspring<sup>11</sup>.

### Dynamic duos

This hints at the possibility of applying the technology to human germline manipulation. But as Chromos only endows its mice with a single chromosome, rather than a pair, only half the offspring end up with a copy. To get artificial chromosomes to inherit normally would require bearers to carry two copies. During the chromosomal dance that takes place in meiosis — the cell divisions that give rise to eggs and sperm — the artificial chromosomes would have to recognize themselves as members of a pair, and segregate accordingly. How this process occurs in natural chromosomes is poorly understood, so there seems to be little chance of making it work in artificial versions any time soon. “We are a very long way from thinking about meiosis,” says Willard.

In any case, therapies that rely on giving people extra chromosomes raise another difficult question. Would people engineered to have 24 pairs of chromosomes instead of the usual 23 be able to reproduce with any partner? Experience with Chromos’s mice suggests that they might produce viable offspring. But unless they mated with another

individual carrying two copies of the extra chromosome, their offspring would only inherit one copy. Their grandchildren, in turn, could not be sure of inheriting the chromosome at all.

Given these complications, some experts believe that the most likely route to human germline manipulation will involve the nuclear transfer technology used to create Dolly the cloned sheep. The basic idea is to take a cell, introduce a transgene by homologous recombination, and then transfer the resulting nucleus to an egg cell stripped of its own chromosomes. Last year, members of the Dolly team at PPL Therapeutics in Roslin, near Edinburgh, created transgenic sheep by this method<sup>12</sup>.

For human germline gene therapy, the procedure would be slightly different. Clinicians would isolate ES cells from embryos made by *in vitro* fertilization (IVF) with the parents’ eggs and sperm. Once the ES cells have multiplied in a dish to a population of several million, a technician would deliver a transgene to all the cells, and select the few in which it had integrated by homologous recombination. Up to this point, the procedure resembles that used to make gene-targeted mice. But rather than making chimaeras, the Dolly procedure would then be applied. The resulting embryo would be genetically identical to the original one created by IVF, with the exception of the modified genes.

**D**oes this amount to sacrificing one life in order to generate another? Or is it the same life all along?

The procedure raises an interesting ethical question. Does it amount to sacrificing one life in order to generate another? Or is it the same life all along? In a sense the original embryo was never destroyed, but recreated by cloning from its own cells.

This ethical conundrum aside, safety is the real barrier. The success rate of nuclear-transfer cloning in animals is still extremely low. Most embryos die before or soon after birth. Even if the miscarriages and infant mortality could be prevented, hidden problems may remain. In July, researchers led by Rudolf Jaenisch at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, reported that mice cloned from ES cells show unpredictably altered levels of gene expression<sup>13</sup>. Until this problem can be solved, experts agree, the procedure cannot be countenanced in humans.

Other biologists argue that there is little point in trying to solve these problems, given the existence of the simpler, safer technique of pre-implantation genetic diagnosis<sup>14</sup>.

### Inheritance taxed

Although it might at first seem sensible to want to repair the single-gene defects that cause diseases such as cystic fibrosis, couples who both carry a mutation in the gene can instead choose IVF, and then have the embryos screened for the mutation. Only those with at least one normal copy are transferred to the mother’s uterus.

In rare cases, pre-implantation genetic diagnosis offers no hope of preventing a heritable disorder. A few Huntington’s disease patients, for example, carry two copies of the faulty gene<sup>15</sup>. A single copy of the Huntington’s gene is enough to cause the fatal neurological disorder, the symptoms of which usually appear during middle age. People with two copies suffer only slightly worse symptoms, but all their children will get at least one copy of the gene.

Such cases are tragedies, but are far too rare to inspire the concerted effort required to develop techniques of germline gene therapy. Any impetus is more likely to come from the discovery of genes that heighten susceptibility to more common diseases. Mutations in the gene *BRCA-1*, for example, can increase the chance that a woman will get breast cancer in her lifetime to 80% (ref. 16).

As the list of known susceptibility genes grows, eliminating them through pre-implantation genetic diagnosis becomes impractical. If a couple carries four disease-susceptibility genes between them, finding an embryo free of all four means screening 16 embryos on average — requiring more eggs than can be taken from a woman in a typical IVF egg-retrieval procedure. On the other hand, the combination of gene targeting and nuclear transfer has the potential — if the problems with miscarriages and infant mor-

## The unregulated frontier



Human germline genetic engineering? Didn't I read somewhere that it had already been done? If that was your initial reaction to this feature, you are probably thinking of the fuss that in March greeted a paper<sup>18</sup> from researchers led by Jacques Cohen at the Institute for Reproductive Medicine and Science of Saint Barnabas in Livingston, New Jersey.

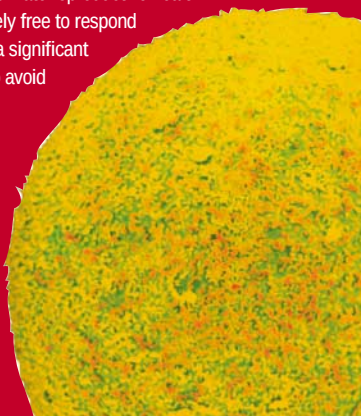
Cohen and his colleagues reported that they had helped infertile women to have children by 'rejuvenating' their eggs with a shot of cytoplasm from a younger woman's egg. Along with the cytoplasm came a handful of mitochondria, the components that generate energy for our cells and which contain a small number of their own genes.

The resulting children each had mitochondria from their mother as well as from the donor, and the authors claimed this amounted to a germline genetic modification.

But this is not what most biologists think of when they use this term. Mixing mitochondria does not risk damaging important chromosomal genes, nor could it endow children with eugenic traits. To call it germline engineering, says Gregory Stock of the University of California, Los Angeles, who has written extensively on the topic, is "playing games with words". Nevertheless, the Cohen paper may provide a preview of how the first genetically modified human child will be born. In some countries, private reproductive health clinics are tightly regulated. But in the United States, they are largely free to respond to the demands of their patients — and surveys have shown that a significant minority of the public is open to the idea of engineering children to avoid disease and even to possess 'desirable' characteristics.

Cohen's work attracted further controversy when it emerged that his paper had failed to mention that two fetuses created by the technique had suffered from a rare chromosomal abnormality — one woman miscarried, the other chose an abortion. To some bioethicists, this underlines the need to regulate private clinics so that they cannot experiment freely with true germline procedures.

**Quick change: a human cell one day after fertilization, right, and a human fetus at seven to eight weeks, top.**



tality can be solved — to repair or replace multiple genes in a single embryo.

But even with safe techniques for making such alterations, critics warn that the genome is far too complex to predict all of

the consequences. Gene variants that contribute to one disease may protect against another — the classic example being the mutation in the gene for haemoglobin that keeps malaria at bay if its bearer carries one copy, but causes sickle-cell anaemia in people unlucky enough to inherit two. Whether a gene variant helps or hurts may depend on its genetic surroundings. In women who are protected from cancer by certain other genes, the risky *BRCA-1* gene might conceivably fend off some other disease.

The problem, say critics, is that we would not know until it was too late that a known disease-susceptibility gene has some beneficial effect. "You need to be careful before you start fiddling around with a delicate system that has evolved over millennia," says Patricia Baird, a geneticist at the University of British Columbia in Vancouver.

But current transgenic animal technology includes methods that would allow the damage to be undone, should problems arise. "It has to be reversible," says Mario Capecchi of the University of Utah in Salt Lake City, one of the pioneers of gene targeting in mice. Reversible transgenes have been used in mice

for years, he points out.

To achieve this in mice, the transgene is placed between the target sequences for a site-specific recombinase, an enzyme that precisely clips out a particular segment of DNA and then reseals the chromosome<sup>17</sup>. In mouse experiments, the recombinase can be activated in specific situations such as the production of sperm cells, allowing the next generation to be bred free of the transgene. For use in human germline gene therapy, the recombinase gene could be included with the transgene and be linked to a regulatory sequence that would be activated if the patient took a specific drug.

### Ethical issues

Although such systems might address the safety objections to germline manipulation, they leave other questions unanswered. Germline gene therapy to replace genes such as *BRCA-1* is seen by some ethicists as being just a short step from manipulations to imbue children with genes predisposing to high intelligence, tall stature or socially desirable behaviour.

Such children might feel like machines manufactured to specifications, says Mary Warnock, a moral philosopher who headed Britain's Committee of Enquiry into Human Fertilisation in the early 1980s — the deliberations of which fed into subsequent legislation on reproductive technology. "If you tinker with the genes of a child to make it turn out any way you want, then there is an intolerable burden on that child," she says.

Clearly, any attempt to manipulate the human germline would provoke a fierce controversy. Those scientists who believe that the technology will one day be perfected agree on one thing: it is far better to hold the necessary ethical debate now, at leisure, before the techniques to make it happen are a practical reality, than later, in haste, when they are entering the clinic.

**Jonathan Knight writes for Nature from San Francisco.**

1. Chan, A. W. S., Chong, K. Y., Martinovich, C., Simerly, C. & Schatten, G. *Science* **291**, 309–312 (2001).
2. Wadman, M. *Nature* **392**, 317 (1998).
3. Frankel, M. S. & Chapman, A. R. *Human Inheritable Genetic Modifications: Assessing Scientific, Religious, and Policy Issues* (American Association for the Advancement of Science, Washington, 2000).
4. Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. & Ruddle, F. H. *Proc. Natl Acad. Sci. USA* **77**, 7380–7384 (1980).
5. Chan, A. W. S., Homan, E. J., Ballou, L. U., Burns, J. C. & Bremel, R. D. *Proc. Natl Acad. Sci. USA* **95**, 14028–14033 (1998).
6. Perry, A. C. F. *et al. Science* **284**, 1180–1183 (1999).
7. Pursel, V. G. *et al. Science* **244**, 1281–1288 (1989).
8. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. & Melton, D. W. *Cell* **56**, 313–321 (1989).
9. Harrington, J. J., Van Bokkelen, G., Mays, R. W., Gustashaw, K. & Willard, H. F. *Nature Genet.* **15**, 345–355 (1997).
10. Telenius, H. *et al. Chromosome Res.* **7**, 3–7 (1999).
11. Co, D. O. *et al. Chromosome Res.* **8**, 183–191 (2000).
12. McCreath, K. J. *et al. Nature* **405**, 1066–1069 (2000).
13. Humpherys, D. *et al. Science* **293**, 95–97 (2001).
14. McLaren, A. *Nature* **392**, 645 (1998).
15. Wexler, N. S. *et al. Nature* **326**, 194–197 (1987).
16. Ford, D. *et al. Lancet* **343**, 692–695 (1994).
17. Chambers, C. A. *Bioessays* **16**, 865–868 (1994).
18. Barritt, J. A., Brenner, C. A., Malter, H. E. & Cohen, J. *Hum. Reprod.* **16**, 513–516 (2001).



**Mary Warnock: worried about effects on children.**



# Brazil has the talent: just let us get on with the job

Frustrated researchers are calling for support from the international scientific community.

Sir — Unlike Europe or the United States, where the demand for people with PhDs is relatively low in the scientific market, Brazil still needs these PhDs in research positions, mainly in the universities, which nowadays are seriously understaffed. Yet we have a dramatic problem, in that these posts are not being filled.

Brazil has a population of nearly 169 million people, about 0.02% of whom have doctorates. A large part of its scientific research is done in the 52 federal universities, supported by the national government.

The country currently lacks teachers and researchers. Areas such as ecology, sociology, economy, history and engineering require applied as well as basic research to address urgent national problems. However, no academic positions have been offered since a governmental decree was passed in November 1997, putting a freeze on both filling and creating research positions. Since then, no new permanent positions have been created, and even those that were already planned have been cancelled, jeopardizing public universities and basic research.

This government policy means that a whole generation of young PhDs will be lost to research. Between 1998 and 1999, for example, only 1,481 new doctors were added to the staff of the country's 52 federal universities, even though 8,790 Brazilian students obtained a PhD during this period.

These young doctors are the result of decades of investment on the part of the Brazilian government; some have obtained PhDs in the world's best scientific centres. Each of these students costs Brazil about US\$22,680 per year if they study abroad, and \$7,250 per year if they study in this country. According to the Brazilian Council for Scientific and Technological Development (CNPq), Brazil spent \$65,631,910 in 1998, and \$38,708,594 in 1999 (a significantly lower investment than the previous year) on grants.

This is the first time that so many well-trained scientists are available to work in Brazil, with the potential to fulfil the needs of our country in various scientific areas. They were intended to enhance Brazil's position in international science, and to open opportunities for collaborative enterprises.

After several protests at the beginning of this year, and long negotiations with the universities, the Brazilian government seemed to be willing to improve the situation. But the position has gone back

and forth since then — on 20 March the government prohibited the creation of any new positions in the federal universities, even temporary contracts. Then, having come under pressure from the federal universities' rectors, the government backtracked and renewed its promise to create 2,000 emergency positions (of the 6,000 needed). But nothing has yet happened.

Brazil's scientific community lacks the political strength to make its protests heard. Therefore, we have decided to mobilize our community to show the government that there is public interest in the development of science in Brazil, and that there is also a qualified workforce available to help our country to overcome its social and economic problems.

## Group did give timely foot-and-mouth analysis

Sir — We write to correct some factual inaccuracies in the Correspondence by R. G. Eddy (*Nature* **412**, 477; 2001), which itself was a response to your Opinion article "Lessons from an epidemic" (*Nature* **411**, 977; 2001).

Eddy suggests that the Imperial College group did not make the details of their research on models of foot-and-mouth virus transmission available to stakeholders until the publication of our paper in *Science* on 12 April. This is incorrect: we sent details of our preliminary analyses to the UK chief veterinary officer on 16 March.

In addition, we were the only group to present — on 23 March — a detailed analysis of the epidemic and potential control options to stakeholders including officials from the UK's Ministry of Agriculture, Fisheries and Food (now the Department for Environment, Food and Rural Affairs), the Office of Science and Technology and the Food Standards Agency. Our presentation included details of the modelling techniques we used.

Since then, we have regularly presented new, updated results to the UK chief scientist's foot-and-mouth disease advisory group (which includes teams from Edinburgh, Cambridge and Pirbright), and have had several discussions with the Royal Society and the National Farmers' Union.

**Roy M. Anderson, Neil M. Ferguson, Christl A. Donnelly**

Readers interested in supporting our efforts are asked to contact S.P.R. by e-mail at [serviopr@mono.icb.ufmg.br](mailto:serviopr@mono.icb.ufmg.br). **Sérvio P. Ribeiro\***, **Milton de S. Mendonça-Júnior†**, **Edésio M. Barbosa‡**, **J. Adauto de Souza Neto§**

\*Departamento de Biologia Geral/ICB, Universidade Federal de Minas Gerais, CP 484 - 30161-970, Belo Horizonte, Minas Gerais, Brazil  
†Laboratório de Ecologia de Insetos, Departamento de Zoologia, UFRGS, Av. Bento Gonçalves, 9500, CEP 91501-970, Porto Alegre, Rio Grande do Sul, Brazil

‡LEM – Laboratoire Environnement et Minéralurgie 15, Avenue du Charmois BP40 F-54501, Vandœuvre-Les-Nancy Cedex, France  
§Departamento de Geologia, Centro de Ciências Exatas e da Terra, UFRN, Lagoa Nova, CEP 59.072-970, Natal, Rio Grande do Norte, Brazil

Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College of Science, Technology and Medicine, London W2 1PG, UK

## To boldly go where no plant has yet been found

Sir — US funding agencies' call for ocean exploration should lead to exciting discoveries, with light being shed on some of the least known portions of our biosphere. But contrary to the details given in the News feature "To boldly go" (*Nature* **412**, 672–673; 2001), scientists involved in the study of Blake Ridge methane seeps have no intention of studying plant life supported by methane at 2,000 metres.

Sunlight penetrates only a few hundred metres into the water column. Although there has been speculation that light generated by thermal radiation from deep-sea hot springs might be sufficient to sustain facultative photoheterotrophic bacteria, the Blake Ridge seeps have no thermal source of light. It would be a surprise, indeed, to discover multicellular plants living in 2,000 metres of water anywhere in the oceans.

**Cindy Lee Van Dover**

Biology Department, College of William & Mary, P. O. Box 8795, Williamsburg, Virginia 23187, USA

**The author intended to say that the researchers would study any life that exists around the Blake Ridge methane seeps. The phrase "plant life" was incorrectly introduced during editing — Correspondence Editor, *Nature***



# The ultimate means of destruction

How the British hydrogen bomb was born.

## Britain and the H-Bomb

by Lorna Arnold

Palgrave: 2001. 273 pp. £45, \$69.95 (hbk),  
£15.99 (pbk)

Stuart Croft

The heights of concern evoked by nuclear weapons during the cold war may now be a memory, but nuclear issues are still important in world politics. India and Pakistan have conducted nuclear tests, apartheid South Africa developed weapons technology, and crises have emerged over North Korea's nuclear capabilities. None of the declared nuclear-weapons states has eliminated all of its weapons, although there have been enormous strides in reducing numbers. Arguments now rage over President George W. Bush's National Missile Defense concept ('Son of Star Wars'). Although they matter less than in the immediate past, nuclear issues do still matter. So, how we arrived at this point is important.

Fortunately, nuclear history is a field of outstanding scholarship. Much has been written about the first two decades of the nuclear age. Most pertinent in the context of Lorna Arnold's *Britain and the H-Bomb* are



Terrible beauty: the deadly mushroom cloud is an enduring symbol of the cold-war years.

two 'official' histories by Margaret Gowing: *Britain and Atomic Energy 1939–1945* (Macmillan, 1964) and *Independence and Deterrence: Britain and Atomic Energy 1945–1952* (Macmillan, 1974), which Arnold herself assisted with. Arnold's new book in many ways carries on where she and Gowing left the story, with access to documents and to personal records that gives the book a remarkable authenticity.

One way of viewing this book is as classic historical biography. Arnold traces the "exciting" days in which the young British H-bomb, the thermonuclear hydrogen bomb, developed and the nature of its career emerged. She does not focus on its maturity, but through a most detailed account of the development of the science and of the tests that demonstrated the technical validity of the weapon, there is a real sense of growth through adolescence.

Many of the protagonists in the story are handled with great care, and one gets as much sense of, for example, Sir William Penney (the head of the nuclear weapons project and director of the UK Atomic Weapons Establishment at Aldermaston, at the time) and his deputy at Aldermaston, Sir William Cook, as in any conventional biography. We are intrigued by the question of who is 'the father of the H-bomb'. And this sense of biography is enhanced by the structure of the book: half of it is concerned with the H-bomb tests on Christmas Island in the Indian Ocean, engrossing the reader in events affecting the development of the young bomb half a world away from Aldermaston.

The scale of the promise of H-bomb technology was dramatic. Britain's first test of an atomic bomb, *Hurricane* in 1952, was 25 kilotons (the world's first test, *Trinity* in the

United States in 1945, yielded 21 kilotons), whereas the *Grapple Y* British H-bomb test in 1958 produced a 3-megaton yield. The Americans tested up to 9.3 megatons. The prospect of the sort of power that the H-bomb promised, over and above that of the A-bomb technology that had devastated Hiroshima and Nagasaki, led Robert Oppenheimer and others who had developed the American A-bomb to describe the new technology as genocidal.

There are, however, some gaps in the analysis. One is strategy. We do not learn a great deal about how the H-bomb was being developed for its career as an 'adult' in deterrence. This is a shame, as those debates were fascinating. Partly, though, the silence on this point is explained by the focus on Aldermaston — strategy did not concern Aldermaston — and partly because the topic has been covered elsewhere, notably in Ian Clark and Nicholas Wheeler's *The British Origins of Nuclear Strategy* (Clarendon, 1989). But a more accurate title, reflecting Arnold's focus on the tests rather than on turning the technology into a deterrent strategy, might perhaps have been 'Testing the technology: Britain's development of the H-Bomb'.

A point of wider interest concerns how little is said, relatively, about two core moral questions. Were the scientists involved, at best, amoral technicians in their development of a weapon of the most terrifying destructive power? And if the aims were amoral, were the means — the nuclear tests themselves — immoral in the impact that they had on human life and on the environment? The first point is touched upon in the analysis: Penney and Cook fervently believed in the doctrine of deterrence. It was their view that these developments should not be



Passionate dissent: not everyone loved the bomb.

HUTTON-DEUTSCH COLLECTION/CORBIS

US DEPT OF ENERGY/SPL

seen in terms of taking the world further towards the brink of nuclear cataclysm; rather, the H-bomb would build a shield of such fearsomeness that war would become inconceivable. For many of those involved, the process of taking the United Kingdom into the H-bomb world was a profoundly constructive act.

The peace movement, emerging as a powerful political force although barely covered in this book, argued precisely the opposite. In 1950, 100 scientists petitioned the British government not to develop an H-bomb. India led a world campaign in favour of a ban on nuclear tests. The Pope condemned the H-bomb. By January 1956, the British prime minister, Anthony Eden, felt compelled to broadcast a message to the nation on the need to keep Britain's options open. And by the mid- and late 1950s, as Christopher Driver detailed in the still-classic *The Disarmers: A Study in Protest* (Hodder & Stoughton, 1964), these were matters of loud and passionate public and private debate; but one would not know that from *Britain and the H-bomb*.

Arnold is perhaps on stronger ground in examining the morality of the means than in justifying the official view of the moral imperative of the need for the H-bomb. Many technical questions are left to a five-part appendix, for those who wish to learn more of the scientific argument. One of these sections focuses on evidence of the health and safety records of the tests. But little is learnt about concern for personnel — weighing the balance of risk against the prospect of developing a technology 'to end wars' — or of the fate of the Christmas Islanders.

*Britain and the H-Bomb* tells the tale of the men who developed Britain's thermonuclear device through the drama of the tests in the Pacific. It does so with clarity and skill. However, some of the wider political and ethical questions still loom large.

Stuart Croft is at the Centre for Studies in Security and Diplomacy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

### More on nuclear weapons

*Fallout: Hedley Marston and the British Bomb Tests in Australia* by Roger Cross (Wakefield Press, 2001; distributed in the United Kingdom by the Airlift book company) gives an Australian perspective on the British A-bomb tests conducted in Australia in the 1950s and their most vociferous opponent there, the fascinating but flawed biochemist Hedley Marston.

*Otto Hahn: Achievement and Responsibility* by Klaus Hoffmann, translated by J. Michael Cole (Springer, 2001), describes the work and thought of one of the discoverers of nuclear fission, and his reflections on scientific and social responsibility.



## Fossilized art

These pictures of fossils from the Triassic era, some 232 million years ago, were painted by Barbara Page and are from *Rock of Ages, Sands of Time* (University of Chicago Press, \$45, £28.50).

With text by Warren Allmon, the book is an illustrated history of the Earth, with each contiguous panel representing one million years, going from the Cambrian to the present.

## Counting on the metaphorical

### Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being

by George Lakoff & Rafael E. Núñez  
Basic Books: 2000. 451 pp. \$30, £21.99 (hbk), \$20, £14.99 (pbk)

Gerald A. Goldin

Modern cognitive science has blossomed in the past few decades. Might its ideas offer us a novel perspective on mathematical thinking, or even on what mathematics itself is?

How do mathematicians reason when they are defining new constructs or exploring abstract ideas? What makes certain ways of thinking rational or logical? If mathematics does not consist of universal truths, what accounts for its remarkable power, apparent timelessness and cross-cultural validity? Why does it describe nature so well? Does it consist of ideas and intuitions, or of theorems and formal proofs? And why do so many students struggle with mathematics, lacking real understanding of why they are manipulating those symbols and formulae?

George Lakoff and Rafael Núñez, linguist and philosopher respectively, answer these questions in a manner surely intended to provoke controversy. They use three main ideas, which they call cognitive science's "recent discoveries about the nature of mind". First, 'the embodiment of mind' is the idea that our bodies and brains, together with our experiences of the everyday world, structure our concepts and reasoning.

Second, 'the cognitive unconscious' is the notion that essential aspects of our thinking, including low-level processes and systems of concept images and relationships, are not accessible to awareness. Finally, 'metaphorical thought' is the idea that we understand abstract concepts concretely in terms of our bodily experiences of sensation and movement, through a mechanism called "conceptual metaphor".

Actually, only the third idea is recent — Lakoff himself, together with Mark Johnson, has developed and vigorously propounded bodily-based conceptual metaphor as a near-universal explanatory construct, accounting not only for the language we use but also for how we think about space, time, life, love, good and bad feelings, and much else. Although it is far from being fully accepted in cognitive science or linguistics, the theory is presented in the book as if firmly established.

With these tools to hand, the authors take apart some of the most important ideas in mathematics. Non-mathematicians will find many of the explanations difficult, but should be able to grasp the general direction of the discussion. Topics covered include symbolic logic, sets and hypersets, transfinite numbers and infinitesimals, fractal curves, Dedekind's construction of the real numbers and Weierstrass's formal definitions in calculus. The book culminates in detailed "case studies" of  $e$  (the base of the natural logarithms), of  $i$  (the square root of  $-1$ ), and of Euler's famous formula  $e^{pi} + 1 = 0$ . For each idea, the authors give us a conceptual metaphor that lies behind the mathematics; that is, imagery that grounds the concept in everyday experience (a "grounding



metaphor”) or that links it to another domain of mathematics (a “linking metaphor”).

Several dozen conceptual metaphors are introduced, and each is given a proper name. Sometimes these are just new names for familiar notions, and add little. Thus, the ‘Measuring Stick’ metaphor lets us associate physical lengths with numbers, and the ‘Numbers are Physical Segments’ metaphor permits the opposite association. According to the “cognitive unconscious” notion, mathematicians typically do not realize that they are thinking metaphorically. For instance, in defining operations on functions with numerical values, they tacitly use the ‘Functions are Numbers’ metaphor. Complicated constructs such as the Cartesian plane (with  $x$  and  $y$  coordinate axes) involve “conceptual blends” of metaphors.

The most interesting and fully developed example of a conceptual metaphor is the ‘Basic Metaphor of Infinity’. Lakoff and Núñez propose this as a general mechanism of cognition that originates outside mathematics, grounded in our everyday experiences with repeated processes (ordinary actions and movements) that come to completion. They call these experiences the source domain. They contend that mathematicians extend these experiences metaphorically to describe “iterative processes that go on and on” — the target domain. Whereas the source domain has a concrete, unique final state, in the target domain the final state is metaphorical and is called “actual infinity”. The metaphor conceptually maintains the uniqueness and finality of actual infinity. By invoking the Basic Metaphor of Infinity, the authors describe as metaphorical a wide variety of mathematical concepts — from proof by induction and transfinite arithmetic, to the symbol  $\infty$  that is used to write formal infinite series.

As a survey of ideas in mathematics, this book does not compare favourably with other popular expositions. Occasional misconceptions, and frequent imprecision of mathematical language in otherwise valid explanations, make close page-by-page reading frustrating. Of course, the authors are not mathematical scientists. Like students new to the subject who are striving to understand the ideas behind formal mathematics, they “discover” that multiplication by  $i$  implements a  $90^\circ$  rotation, that space-filling curves do not fill space (when hyper-real coordinates are included) and that symbolic logic is “not absolutely true”. Such interpretations, while not original, are the best parts of the book. But Lakoff and Núñez write as if they are the first to discover them, calling them “not new mathematical results, but new ways of understanding well-known results”. They seem unaware that these and similar ideas are commonly used by good teachers of mathematics. And they often seem to assume, quite unjustifiably, that each mathematical construct can

be understood in only one such way — the one they have discovered — and that they have found the *real* metaphor from which the mathematics originates.

Lakoff and Núñez make many far-reaching claims based on their conceptual-metaphor analysis. They regard their work not merely as explaining imagery in mathematical thinking, but as profoundly affecting mathematics itself. They claim to have overturned what they call “The Romance of Mathematics” — conventional beliefs they attribute to most mathematicians, such as “mathematical truth is universal, absolute, and certain”, and “the book of nature is written in mathematics” — and to have sketched, for the first time, what mathematics really is. They see their philosophy of “mind-based” or “embodied” mathematics as inconsistent with any existing philosophy, and “mathematical idea analysis” as a new discipline they themselves have launched, dethroning the queen of the sciences. Mathematicians sceptical of their ideas, they suggest, are likely to be Platonists (who believe in ideal forms), naive realists or empty formalists, influenced by self-interest, elitism and possibly a sense of wounded identity.

The arguments they offer as to why mathematics is not absolute are mostly familiar, although presented as novel. Their philosophical direction, which relies on our having access only to mathematics developed through human cognition, resembles well-known arguments for the essential subjectivity of all human knowledge. And the arguments are directed against naive, stereotypical opinions; serious philosophers are neither quoted extensively nor challenged. The profound evolution of mathematicians’ understanding of ‘truth’ and ‘existence’, stretching across millennia, is omitted entirely.

I was most troubled by the narrowness of the book’s base in cognitive science. No relation is acknowledged to most other work involving sensorimotor experiences, concrete and abstract understanding, or imagery in mathematics. There is some discussion of the brain, of arithmetic in animals and human infants, and of schemas and cognitive operations, but almost nothing about learning, human developmental and cognitive psychology, or problem-solving heuristics and strategies. Essential ideas of cognitive science, such as analogical reasoning, systems of internal representation, information-processing models, developmental stages, cognitive structures, or affect and motivation, find no place. Rather, “conceptual metaphor” subsumes or excludes all other constructs. By treating everything as metaphor — mathematical statements, definitions, proofs, representations and models, as well as notations, images, analogies, generalizations, mappings, conceptualizations and examples — Lakoff and Núñez disregard cognitively

important distinctions. And the notion of metaphor itself loses explanatory power.

In short, although I strongly favour analysing the ideas and imagery of mathematics, and agree with the authors’ view that the “portrait of mathematics” has a human face, I regard this book as fundamentally flawed. Its flaws are independent of the straw opponents its authors set up. ■

Gerald A. Goldin is in the Departments of Mathematics and of Physics and Astronomy, Rutgers University, New Brunswick, New Jersey 08903, USA.

## Sticking by our one and only?

**The Myth of Monogamy: Fidelity and Infidelity in Animals and People**

by David P. Barash & Judith Eve Lipton  
Freeman: 2001. 288 pp. \$24.95, £18.99

T. R. Birkhead

In the mid-1970s, David Barash was one of the first biologists to embrace the new science of sociobiology and to look at adultery from an evolutionary perspective. He studied bluebirds, which, like most birds, were assumed to be monogamous. Barash wanted to know whether fear of female infidelity and the potential loss of paternity shaped



Birds do it, bees do it — even blue-footed boobies do it. But do they do it monogamously?

BBC WILD



male behaviour. So he placed a stuffed male bluebird in the territory of a breeding pair at different stages of their reproductive cycle. His aim was to test the idea that, to protect his paternity, the male would respond to the 'intruder' more aggressively when his partner was fertile than at any other time.

That is what Barash found, albeit on the basis of just two pairs of birds, and his results were subsequently published in the prestigious journal *American Naturalist*. Barash's approach and his conclusion that the results were "consistent with evolutionary theory" incensed critics of sociobiology. The evolutionary biologist Stephen J. Gould used Barash as a whipping-boy, asserting that he had produced an untested and untestable 'just-so' story.

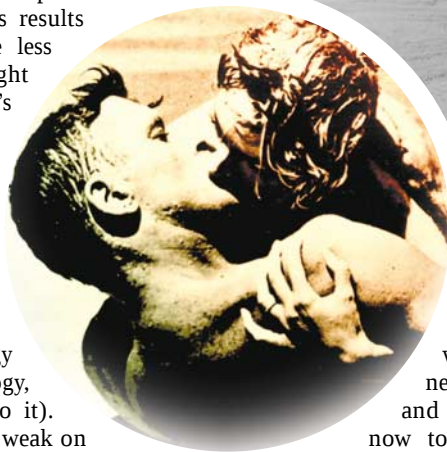
No doubt the trouble partly arose because of where Barash published his findings. Had his results appeared somewhere less influential they might have escaped Gould's disapprobation. Another of Barash's studies, this time on 'rape' in mallard ducks and published in *Science*, also received much criticism both within and outside sociobiology (or behavioural ecology, as Europeans refer to it). Barash's studies were weak on data but strong on ideas — not an uncommon phenomenon in a new area of research as sociobiology was then, and a recent reappraisal of these controversial papers (in *The Triumph of Sociobiology* by John Alcock; Oxford University Press, 2001) exonerates him and commends his use of the scientific method in these studies.

Barash was a pioneer — one of the prime movers in a field that subsequently blossomed into the study of mating systems, including monogamy and its deviations. After a brief flurry of innovative papers in the 1970s, however, Barash disappeared from the primary literature. But now, 25 years and several books later, he has re-emerged with *The Myth of Monogamy*, co-authored with his partner Judith Eve Lipton, a psychiatrist. They have produced a highly readable, light-hearted survey of monogamy and its variations across the animal kingdom, including humans.

The evidence, thanks largely to molecular paternity analyses — DNA testing, in common parlance — is absolutely clear: females of most animal species, including ourselves, routinely copulate with more than one male. Although humans and most



**Classic adultery:** Burt Lancaster and Deborah Kerr on the beach in the film *From Here to Eternity*.



birds appear to have monogamous pair bonds, the molecular evidence tells us that what you see is not necessarily what you get, and biologists are careful now to distinguish between social monogamy and genetic monogamy.

Sex sells, so they say, and human sex cells are at the heart of this story, which focuses on the question of whether a female's ova are fertilized by her partner's sperm or someone else's. No doubt this book will titillate, romping as it does through the vast diversity of human marriage arrangements, although making sense of human mating systems is far from easy. While *The Myth of Monogamy* assiduously covers the extensive literature on this topic, the examples used often seem to be those that provide an appealing story rather than those whose results have been confirmed (or refuted) by subsequent research. This probably doesn't matter too much, for the take-home message is simple: monogamy, as most people know it, is unnatural. We are naturally promiscuous — male and female both — but monogamy is a cultural artefact imposed upon many of us in the Western world by religious rules in an effort to even out the differences between individuals and to stabilize society.

Barash and Lipton discuss the behaviours of human males that are designed to protect their paternity. They argue that behaviour such as jealousy, which often results in the

punishment of both unfaithful wives and their lovers, would be unnecessary if female infidelity had not been part of our evolutionary history. Evolutionary psychologists often use universal patterns of behaviour across cultures as evidence that particular traits have a genetic basis. By showing that men have similar responses to infidelity across cultures, Barash and Lipton consolidate this view.

Nonetheless, some curious puzzles remain. In some cultures, female adultery is not punished; indeed, it can seem to be actively encouraged. In certain Inuit societies, men offer their wives to guests as 'hospitality', whereas in others a wronged husband merely challenges the lover to a singing contest! These exceptions require explanation, and a closer look at the ecological circumstances in which these different cultures have evolved might have provided some clues.

For many people, monogamy and morality are synonymous, and some readers may feel that Barash and Lipton's assertion that monogamy is unnatural is an invitation to behave promiscuously. But the authors are extremely careful in traversing this political minefield, pointing out that there is little connection between what is natural and what is 'good'. They conclude reassuringly by telling us that although monogamy may not be easy — because it isn't natural — it is well within the realm of human possibility. ■

*T. R. Birkhead is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.*

# A tale from Bioutopia

Could a change of nomenclature bring peace to biology's warring tribes?

Pier Luigi Nimis

Once upon a time, two tribes dominated Bioutopia. The small but powerful tribe of Real Taxonomists occupied several scattered ivory towers in the mountains. The huge but poor tribe of Name-users lived in the swamps. They both worshipped Names, but with different rites. The Name-users peacefully adored a huge book made of granite, in which billions of Names were inscribed for Eternity. The favourite occupation of the cruel Real Taxonomists was sacrificing a few Names every day, just by changing them. This they did after consulting their Oracle, Phylogenia, who lived in a cloudy forest.

One day, Phylogenia started to worship a new God, called DNA, and uttered the following words: "Why should *Lichenia splendens* stay together with *L. tristis* and those other 347 species in the same genus? DNA has spoken. The Names must change!" None of the Name-users had complained when *L. tristis* was transferred from the family Licheniaceae to the Tristidaceae, but they all got upset when the Real Taxonomists decided that these 348 species must be called *Thundertenthronkia*, because DNA had spoken. The trouble started at a meeting of the Parliament of Bioutopia. They had to change the law protecting *L. tristis*, the official state organism, and they refused to rename it *Thundertenthronkia tristis*.

Then the war started. Billions of Name-users on one side, the few Real Taxonomists on the other. A fire was burning all around that cruel battle, the same fire that all of us quietly host in the warm shelter of our beloved binomial system.

Generic epithets are indeed like viruses. They are carriers of dangerous phylogenetic implications that kill names. As a sin of my old age, three years ago I tried to warn against the rash acceptance of new generic names. I was so naive as to suggest that the Real Taxonomists should consider the needs of the poor Name-users.

When I reread that article recently, I saw myself as an old Victorian lady fighting against the outrageous trend of wearing skirts so short as to expose the knee. Now, after three years of molecular brainstorming, even the nuns of Bioutopia go around in miniskirts. The thin umbrellas of Victorian ladies cannot fight against the hurricane of generic changes that is ahead. And why should they try? Why should one fight against something fresh, exciting and so

scientifically sound? Those who worship books of granite cannot hinder a free development of (r)evolutionary taxonomy.

There is a sentence engraved on the stone cover of the Name-users' book: "Nomina si nescis, perit et cognitio rerum", which means: "If you do not know the Names, Knowledge is also dead for you." The Name-users explained to me that humans, the only animal to develop language, cannot worship a dictionary from which 10% of the names are scraped out every year. This made me think. Name-users gain knowledge by learning and using names. But the Real Taxonomists produce brand new knowledge for mankind. Why should these tribes fight against each other? Do we really need this conflict? Do we really need generic names?

I therefore consulted another Oracle, called Logic, and she dictated: "Get rid of the binomial." Every species in Bioutopia could be designated by a single epithet: a number, or a barcode, the best food for computers. Surely Name-users — such as curators of collections and databases, authors of books and identification keys, legislators and teachers — should be happy with something like: "It's an *X157YR22297*!" The Real Taxonomists could then concentrate on more important matters than scraping Names off granite, and Phylogenia would be free to change her mind whenever she liked. The Name would remain the same. Peace would return to Bioutopia.

Yet I wonder how many amateurs, having found a rare *L. splendens* on an old oak, will exclaim joyfully to their companions: "Wow! Look at this! I've found *X157YR22297*!" I also wonder what they should say when, more cautiously, they think they've found just "a *Lichenia*". Surely not "I've found something starting with *X157*..."

Last week I read an article about the pressure from public authorities in the United Kingdom to create vernacular names for organisms. That article has come to my mind now, but in association with a different question: "Is *L. splendens* better

Humans cannot worship a dictionary from which 10% of the names are scraped out every year.



**Tower of power: the exalted Real Taxonomists, armed with knowledge, can change the Names.**

than *X157YR22297* for an amateur who reads only Chinese?" Perhaps *L. splendens*, although not a 'vernacular' name, could find a place in the list of 'mid-level names' — those that exist for half-educated people, like most of us, who would have problems in remembering what *X157YR22297* is.

As a citizen of Bioutopia, I have an identity card. It bears my social security number (NMSPLGP09etc., I always forget it) plus my Name (Pier Luigi Nimis), although my parents call me "Pil" (my vernacular name). I do not see anything wrong in being named 'NMSPLGP09etc.' in all official transactions with my government. If this works for people, why should it not work for organisms?

There is an alternative solution: let things continue the way they're going. The war will eventually stop when every single species belongs to a monospecific genus. ■

Pier Luigi Nimis is in the Department of Biology, University of Trieste, Via L. Giorgieri 10, I-34127 Trieste, Italy, and is president of the International Association of Lichenology.

THE CASTLE BY MARK BARING. PRIVATE COLLECTION/BRIDGEMAN ART LIBRARY



# The self 'out there'

Igor Aleksander

Despite sounding like the mother of all oxymorons, the concept of a conscious machine is gaining credibility. Certainly, no actual machine could be described as being conscious. But neurologists and engineers are developing ever more accurate models of the specific chemical and electrical activity in living brains that is necessary for a specific conscious sensation to arise. As this understanding is mechanistic, it raises the serious possibility that an implementation of such models could give rise to machines that are driven by the same mechanisms, and hence are 'conscious' in a non-biological way.

What features does consciousness have in

living organisms? First, to be even minimally conscious, any organism must be aware of its presence in an external world with which it can interact with respect to its needs. This favours the creation (by evolution or design) of successful mechanisms that support this interaction. Few would deny that in natural organisms this sensation is uniquely tied to the electrochemical activity of groups of neurons, which in higher organisms form a brain. This activity is unique in the sense that two distinct sensations cannot be due to the same brain activity without evoking ghostly intermediaries.

Second, what we imagine and recall involves resonances between neural layers, stimulating activity that originated during perception. So imagination 'feels' like a recall of perception even if the exact perception had never taken place. What makes consciousness so puzzling is that the neural activity responsible for both perception and imagination provides a sensation of being an observer (the self) in an 'out-there' world. This sensation depends not only on sensory neural activity but also on virtually unconscious exploratory motor neural activity due to conscious curiosity or to unconscious habit, instinct or reflex.

Take the oculomotor system, for example, which has evolved to allow the eyes to move rapidly towards tiny changes in the field of view, to follow moving objects, to converge if something comes nearer, and even to move rapidly towards a perceived sound. It also interacts with memory to check hypotheses about partially seen objects and to predict their unseen parts. So the neural activity that supports conscious sensation not only involves sensory signals such as those generated by incident light on the retina or vibratory stimulation of the cochlea, but is clearly dependent on signals from the parts of the body that move the eyes and the head, and signals from touch. There is much evidence in neurology (for example in the work of Carlo Galletti, which began in 1989) that cells in various cortical areas only process sensory information as indexed by muscular action, creating inner representations of events that take place 'out there'. Some of this supporting unconscious neural activity is what Christof Koch and Francis Crick call "the zombie within".

The key step in accepting that a machine can be conscious is to realize that when humans describe their sense of consciousness (including its strong qualitative content, sometimes called 'qualia'), they are describing neural activity that has 'out-there' properties, and that there is no real barrier to machines doing the same. Such thinking is

## Artificial consciousness

*Are we nearing a time when sensation can be mechanistically explained outside biology?*

reflected in philosophy too: 'out-there-ness' is the term used by Max Velmans to describe the 'reflexive' nature of consciousness. Work by Jack Cowan and colleagues at the University of Chicago has shown that appropriate computer models of the human visual system can give accurate predictions of the hallucinatory sensations reported by drug users. This puts paid to criticisms that sensation cannot be mechanistically explained except through biology.

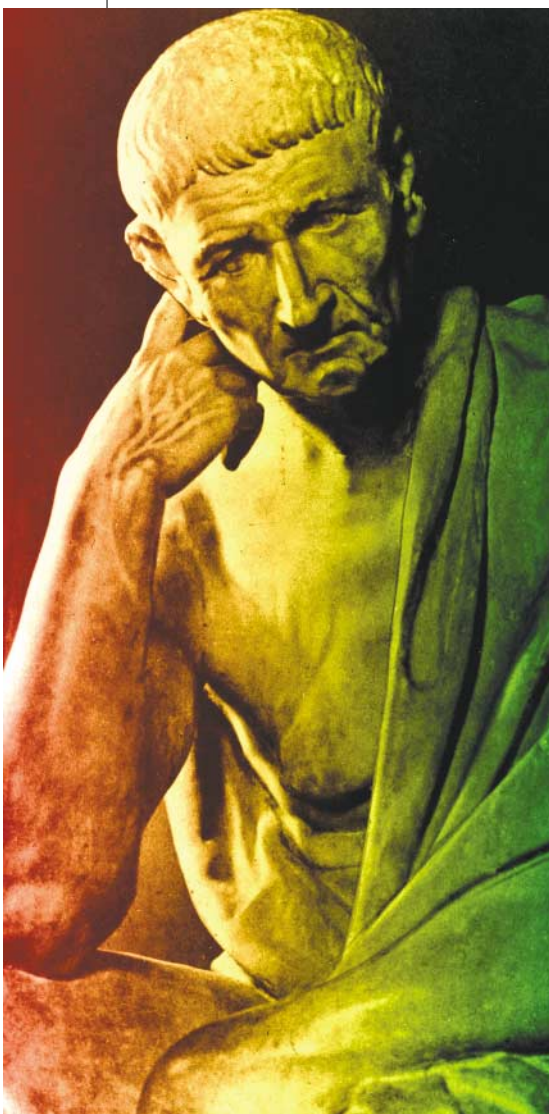
Another criticism is that even if 'out-there' mechanisms were transferred into a robot, it would only become a well-behaved, unconscious zombie because it would still lack 'ingredient X', which turns the zombie into a conscious organism. The conscious-machine concept calls for a fair argument. The machine constructor will attempt to demonstrate that ingredient X is not necessary, whereas the detractor will have to prove that it is, which has not yet been done.

Of course, any robot constructed with the features I have described might be able to perceive and imagine itself in a visual world. More speculatively, emotion, desire, ambition, joy and depression, which also have a neural basis, would become candidates for being transferred into engineered artefacts. As far as the conscious robot goes, it is not our emotion, desire and so on that would be transferred to it — rather, it would have a non-biological neural structure, only developing emotions that would be appropriate to its own existence. It would share with living beings the evolutionary, emergent, depictive and interactive mechanisms that make us conscious. ■

*Igor Aleksander is in the Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.*

### FURTHER READING

- Aleksander, I. *How to Build a Mind* (Phoenix, London, 2001).  
 Aleksander, I. & Dunmall, B. *Proc. R. Soc. Lond. B* **267**, 197–200 (2000).  
 Bressloff, P. *et al. Phil. Trans. R. Soc. Lond. B* **356**, 1–32 (2001).  
 Koch, C. & Crick, F. *Nature* **411**, 893 (2001).  
 Galletti, C. & Battaglini, P. P. *J. Neurosci.* **6**, 1112–1125 (1989).  
 Velmans, M. *Understanding Consciousness* (Routledge, London, 2000).



**Out-there philosophy:** Aristotle stressed the crucial role of sensory input in assembling a conscious appreciation of the outside world.

# X-rays from the edge of infinity

Fulvio Melia

The supermassive black hole at the centre of our Galaxy has a strong influence on its surroundings. Astronomers cannot yet see this beast directly but they now have a much better idea of its size.

The centre of the Milky Way lies behind a dusty cloud of gas and a clustering of stars in the constellation Sagittarius. The closest object to the actual centre, discovered in 1974 (ref. 1), is a bright compact source of radio waves known as Sagittarius A\*. Some ten million stars swarm within a light year of its position, the closest three or four of which swing past it at incredible speeds, up to five million kilometres per hour<sup>2</sup>. The central mass required to keep these stars in their frantic orbits (just 7 light days away from the centre) is equal to 2.6 million solar masses, which suggests that the dark force powering Sagittarius A\* is a black hole — an object so dense that gravity prevents everything, even light, from escaping. All this matter appears to be concentrated at the Galactic Centre in a region no bigger than our Solar System, and probably smaller than the region within Mars' orbit around the Sun<sup>3</sup>.

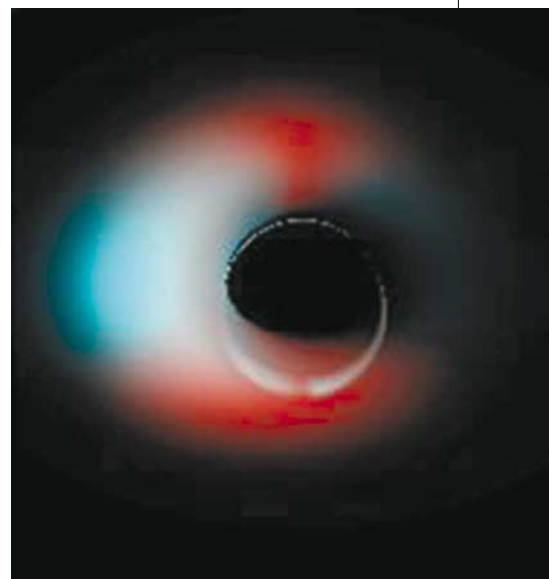
Despite this compelling, though indirect, evidence for a supermassive black hole at the centre of the Milky Way, some astronomers wonder whether it is a true black hole as predicted by Einstein's theory of general relativity. If Sagittarius A\* is a black hole, rather than some other condensation of dark matter, such as a clump of superdense dark stars, then all of its mass must be contained within its event horizon — the virtual membrane that forever separates the world within it from the Universe in which we live. But the orbits traced by the stars closest to Sagittarius A\* are approximately 30,000 times larger than the event horizon predicted for a black hole of 2.6 million solar masses, so we cannot completely rule out alternative explanations. To confirm its existence beyond any doubt, astrophysicists would like to either 'see' the black hole directly<sup>4,5</sup>, or detect an activity occurring barely outside it. On page 45 of this issue, Baganoff *et al.*<sup>6</sup> offer the best evidence yet — they report the rare detection of an X-ray flare occurring just outside the point of no return.

Nature must surely be a chiaroscuroist, for in its cosmic canvas we can discern one of the greatest paradoxes in science — that 'black' holes tend to be among the brightest objects in the Universe. This happens because strong sources of gravity, such as Sagittarius A\*, do not exist in isolation. Rather, their presence is felt with overwhelming force by any matter and radiation wandering haplessly nearby. Sagittarius A\* shines by proxy, inducing the

environment to radiate on its behalf, thereby revealing its presence indirectly (Fig. 1). The recently launched Chandra X-ray Observatory has been sensing this radiation on and off since 1999, producing X-ray images of Sagittarius A\* and its nearby environment with an unprecedented resolution of about 0.5 arcseconds, corresponding to roughly one-twentieth of a light year. Baganoff *et al.* argue convincingly that these X-rays must be coming directly from Sagittarius A\*, rather than a more mundane object, such as a binary star, given the low probability of this chance coincidence. Still, nature has been cruel to astronomers in other circumstances, and this probability, although small, is not zero.

The event reported by Baganoff *et al.* is the remarkable discovery by the Chandra X-ray Observatory, late last year, of a short-duration, highly variable, X-ray flare produced by Sagittarius A\*. It was over in a couple of hours, but while it lasted about 45 times as many X-rays were being emitted per second as are usually produced when Sagittarius A\* is quiet. Even more startling was the observation that, during the flare, the X-ray output dropped abruptly by a factor of about five in less than ten minutes, and then recovered almost as quickly. Such breathless variability is rarely seen in emissions from ponderous multi-million solar-mass objects. It is far more common, however, with solar-sized objects, such as a neutron star accreting gas from its orbiting companion. And that's what makes it interesting.

Variability is a powerful indicator of the size of the radiating region, because nothing can travel faster than light. For an entire object's luminous output to vary considerably over a short interval, all its various parts must be able to 'communicate' their changes to each other within that time. An abrupt change in the X-ray emission from Sagittarius A\* over ten minutes means that the compressed, hot, radiating gas could not have been distributed over a region bigger than the distance between Earth and the Sun, this being the distance traversed by a light beam during that interval. By comparison, the radius of Sagittarius A\*'s event horizon is thought to be only about 20 times smaller. This is why the discovery reported by Baganoff *et al.* is so exciting, because it places the 2.6 million solar masses calculated from the movements of stars into a region at most 20 times larger than that predicted for this



**Figure 1 Hovering on the precipice.** A simulation of what the supermassive black hole at the centre of our Galaxy might look like<sup>7</sup> if you were looking down at the hot, compressed gas swirling around its event horizon. The cyan colour indicates the emission of radio waves oscillating sideways (horizontal polarization), whereas red indicates waves oscillating vertically. The bright, thin ring is due to light that has circled the black hole once before leaving its environment. Baganoff *et al.*<sup>6</sup> report that the Chandra X-ray Observatory detected a short-lived X-ray flare produced by this hot plasma in October last year. The discovery of this activity, located just outside the predicted event horizon, suggests that a black hole is the only plausible explanation for the condensation of dark matter at the centre of the Milky Way.

black hole by general relativity. By squeezing the radius allowed for the central dark matter by a factor of 1,500, Baganoff *et al.* appear to have excluded all other plausible alternatives.

But how can we be sure that we are not the victims of a mischievous hoax by nature? Fortunately, Sagittarius A\* is also detectable at radio wavelengths, and Baganoff *et al.* report that, during Chandra's observation last year, the radio emission from Sagittarius A\* was also higher than normal. This observation increases the likelihood of the X-ray flare being associated with the black hole. Incontrovertible evidence that the X-ray variability is unmistakably from Sagittarius A\* will come if and when we see simultaneous



flaring at both X-ray and radio wavelengths<sup>3</sup>. This will be the icing on the cake because Sagittarius A\*'s radio emission is unique in the Galaxy, and is definitely not produced by an intervening binary system. Coordinated observations such as this are now being planned, and we may have a definitive answer within a year. ■

Fulvio Melia is in the Department of Physics and Steward Observatory, The University of Arizona, Tucson, Arizona 85721, USA.

e-mail: melia@physics.arizona.edu

1. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
2. Ghez, A. M., Morris, M., Becklin, E. E., Tanner, A. & Kremenek, T. *Nature* **407**, 349–351 (2000).
3. Melia, F. & Falcke, H. *Annu. Rev. Astron. Astrophys.* (in the press).
4. Falcke, H., Melia, F. & Agol, E. *Astrophys. J.* **528**, L13–L17 (2000).
5. Cash, W., Shipley, A., Osterman, S. & Joy, M. *Nature* **407**, 160–162 (2000).
6. Baganoff, F. K. et al. *Nature* **413**, 45–48 (2001).
7. Bromley, B. C., Melia, F. & Liu, S. *Astrophys. J.* **555**, L83–L87 (2001).

## Cardiovascular biology

# Platelets and proteases

Skip Brass

Circulating platelets are essential for the formation of blood clots. Studies of mice reveal more about the proteins involved in activating platelets, with implications for understanding strokes and heart attacks.

How do animals stop bleeding after an injury? By the time humans evolved, a common solution lay in rapidly generating a complex clot made up of blood-derived proteins (fibrin) and cells (platelets). Working together, fibrin and platelets create a plug that prevents bleeding long enough for healing to occur. The downside is that clots can also form in blood vessels scarred by cholesterol-laden plaques (atherosclerosis), interrupt-

ing blood flow and leading to strokes and heart attacks.

Writing on page 74 of this issue, Sambrano and colleagues<sup>1</sup> investigate one of the key unanswered questions about clotting: what is the role in mice of a protein called protease-activated receptor-4 (PAR4)? Their results confirm the need for this protein in clotting, but the implications go beyond this, touching on the fine-tuning of platelet activation and intriguing differences between mice and

humans. There may also be implications for developing drugs to prevent cardiovascular problems.

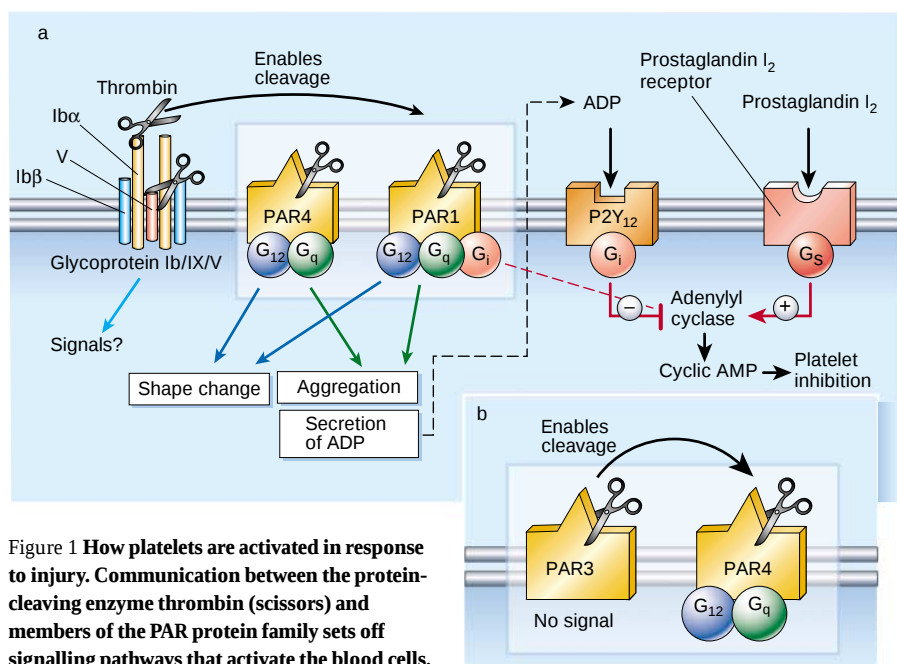
One of the earliest steps in the formation of a blood clot is the localized generation of the active enzyme thrombin from its circulating inactive precursor, prothrombin. Thrombin cleaves fibrinogen in blood plasma, creating monomers of fibrin that then polymerize into a mesh across a wound in a blood-vessel wall. Thrombin also activates platelets<sup>2</sup>. A study published a decade ago<sup>3</sup> established that human platelets express on their surface protease-activated receptor-1 (PAR1), a signalling protein that is a substrate for thrombin. By cleaving PAR1, thrombin initiates signalling within platelets, causing them to stick to each other. The basic structure of PAR1 places it within a large family of cell-surface receptors that work through intracellular switches called G proteins.

With PAR1 identified, the mystery of how a plasma protease could activate platelets seemed to have been solved. But before long, questions began to accumulate. If PAR1 is indeed the universal thrombin receptor, why did deleting its gene in mice have no effect on platelet activation by thrombin? How can cleavage of a single type of receptor activate more than one signalling pathway within platelets, and how could these responses be modulated over time and over the wide range of thrombin concentrations that platelets are likely to encounter? What, if any, is the role of other thrombin-binding proteins found on the platelet surface?

Answers to at least some of these questions are now known. As it turns out, there are three other members of the PAR family<sup>4</sup>: PAR2, PAR3 and PAR4. Like PAR1, two of these receptors — PAR3 and PAR4 — are activated by thrombin; PAR2 is not. Human platelets express PAR1 and PAR4, whereas mouse platelets express PAR3 and PAR4 but not PAR1. This, of course, explains why the loss of PAR1 had no effect on platelet function in mice.

Thrombin produces a variety of responses in platelets because the PAR receptors are coupled to more than one type of G protein. Graded responses to a range of thrombin concentrations can be explained in part by a relationship between the amount of thrombin present and the number of PAR proteins cleaved per second, and in part by differences among PAR family members in the concentration of thrombin required for cleavage. So, for example, PAR1 is a better substrate for thrombin than PAR4. When both are present, cleavage of PAR1 presumably precedes cleavage of PAR4 as thrombin accumulates (Fig. 1a)<sup>5,6</sup>. On the other hand, once PAR4 is cleaved, it appears to signal for longer<sup>7</sup>.

The fact that human platelets express PAR1 and PAR4 whereas mouse platelets express PAR3 and PAR4 seems to be an



**Figure 1** How platelets are activated in response to injury. Communication between the protein-cleaving enzyme thrombin (scissors) and members of the PAR protein family sets off signalling pathways that activate the blood cells.

a. In humans, platelets express PAR1 and PAR4. These are coupled to intracellular signalling pathways through molecular switches from the G<sub>q</sub>, G<sub>12</sub> and G<sub>i</sub> protein families. When thrombin removes the amino termini of PAR1 and PAR4, several signalling pathways (coloured arrows) are activated, one result of which is the secretion of ADP. By binding to its receptor, P2Y<sub>12</sub>, ADP activates additional G<sub>i</sub>-mediated pathways. In the absence of wounding, platelet activation is counteracted by signalling from prostaglandin I<sub>2</sub>. Cleavage of PAR1 by thrombin appears to be enabled by thrombin binding to glycoprotein Ib $\alpha$ . b. Mouse platelets express PAR3 and PAR4. PAR4 seems to be solely responsible for signalling<sup>1</sup>, whereas PAR3 enables cleavage of PAR4, allowing it to respond to lower concentrations of thrombin.

example of divergent solutions to a common problem — how to induce a variety of responses to thrombin. Although mouse PAR3 is a good substrate for thrombin, it signals poorly, if at all. In other words, it has for some reason lost the ability to activate G proteins, while retaining the ability to be cleaved by thrombin. Nonetheless, when the gene encoding PAR3 was deleted from mice, platelets responded considerably less well to thrombin<sup>5</sup>. So it was suggested that PAR4 is the main mediator of mouse responses to thrombin, and that PAR3 has a supportive role, enabling the cleavage of PAR4 at low thrombin concentrations. If so, then one would predict that deleting PAR4 would abolish thrombin signalling in mouse platelets.

Sambrano *et al.*<sup>1</sup> have now deleted the PAR4 gene in mice, and find that platelets are indeed no longer activated by thrombin. The results show the requirement for PAR4 and support the predicted helper role for PAR3 (Fig. 1b). For this awkward arrangement to have arisen, mouse PAR3 not only had to lose its intrinsic ability to signal when cleaved; it also had to 'acquire' the ability to aid PAR4 activation. One might speculate that these differences between mouse and human platelets are not mere happenstance, but are appropriate for other reasons. That remains to be seen.

As the authors show, platelet responses to thrombin were completely abolished in mice lacking both copies of the PAR4 gene. Mice with one copy of the gene showed a modest impairment in clot formation. Although this impairment did not reach statistical significance, there seems to be a trend here: mouse platelets may require that most, if not all, of their thrombin receptors are fully functional. If this is true of human platelets, too, then inherited variations in the structure of PAR family members that result in a reduction in signalling might translate into differences between people in their risk of developing strokes and heart attacks. So far, no function-altering variations in PAR1 and PAR4 have been reported, but one has been described<sup>8</sup> for PAR2.

Are the PAR proteins sufficient to account for all of the responses of platelets to thrombin? Given the new results<sup>1</sup> and the fact that simultaneous blockade of PAR1 and PAR4 abolishes the ability of human platelets to respond to thrombin<sup>9</sup>, the answer would appear to be 'yes'. But recent studies<sup>10,11</sup> have reinvigorated a debate about the role of the glycoprotein Ib/IX/V multiprotein complex as a binding site and substrate for thrombin. It is still too early to predict the outcome of that debate. Nonetheless, the results from Sambrano *et al.* provide a conclusive demonstration of the essential role of PAR4 in the activation of platelets by thrombin and, by extension, of the need for thrombin itself. Their work also re-emphasizes the possibility that drugs that block PAR1, PAR4 or both

might be useful in treating a variety of clotting disorders in humans.

*Skip Brass is in the Departments of Medicine, Pathology and Pharmacology, and the Center for Experimental Therapeutics, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, Pennsylvania 19004, USA.*

*e-mail: brass@mail.med.upenn.edu*

1. Sambrano, G. R., Weiss, E. J., Zheng, Y.-W., Huang, W. & Coughlin, S. R. *Nature* **413**, 74–78 (2001).
2. Niewiarowski, S. & Thomas, D. P. *Nature* **212**, 1544–1547 (1966).

3. Vu, T.-K. H., Hung, D. T., Wheaton, V. I. & Coughlin, S. R. *Cell* **64**, 1057–1068 (1991).
4. O'Brien, P. J., Molino, M., Kahn, M. & Brass, L. F. *Oncogene* **20**, 1570–1581 (2001).
5. Kahn, M. L. *et al.* *Nature* **394**, 690–694 (1998).
6. Xu, W.-F. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 6642–6646 (1998).
7. Covic, L., Gresser, A. L. & Kulipulos, A. *Biochemistry* **39**, 5458–5467 (2000).
8. Compton, S. J. *et al.* *J. Biol. Chem.* **275**, 39207–39212 (2000).
9. Kahn, M. L., Nakanishi-Matsui, M., Shapiro, M. J., Ishihara, H. & Coughlin, S. R. *J. Clin. Invest.* **103**, 879–887 (1999).
10. Ramakrishnan, V. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 1823–1828 (2001).
11. Dörmann, D., Clemetson, K. J. & Kehrel, B. E. *Blood* **96**, 2469–2478 (2000).

## Earth science

# Core beliefs

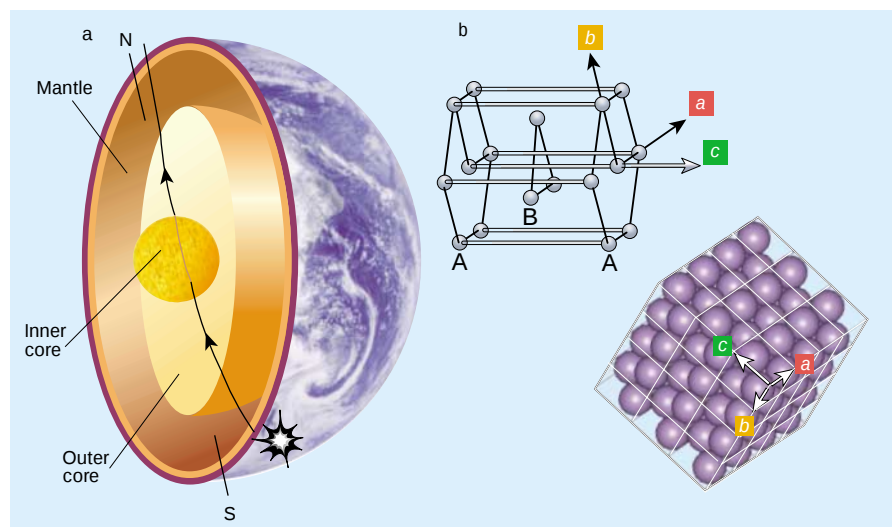
Andrew Jephcoat and Keith Refson

Working out what happens in the extreme conditions at the centre of the Earth is not easy. A calculation of the properties of iron under such conditions helps to explain seismic observations of the inner core.

In the beginning, shortly after the Earth was formed, there was no inner core. This has since grown to a region of dense, nearly pure solid iron, some 2,440 km across. We know its size and something of its internal structure because of its effect on seismic waves produced by earthquakes, which pass through the core on their way to detector stations at the surface (Fig. 1a). A longstanding puzzle is that the speed of a seismic compressional wave (an ultralow-frequency sound wave) depends on its direction across

the core. Such waves travel faster along the axis of Earth rotation (north–south) than in the equatorial plane (east–west). This difference — seismic anisotropy — appears to increase with depth within the inner core and is patchy on a variety of length scales<sup>1–3</sup>. Other vital evidence for seismic anisotropy comes from free oscillations — studies of the natural vibrational frequencies of the Earth.

On page 57 of this issue Steinle-Neumann *et al.*<sup>4</sup> present calculations of the behaviour of iron at high temperatures and pressures. They



**Figure 1 Studying the inner core.** a, Cross-section through the Earth showing a typical path taken by seismic waves through the inner core. The star marks the approximate location of an earthquake event on the South Sandwich Islands. (Earth image taken by Galileo orbiter, modified from <http://photojournal.jpl.nasa.gov>) b, The hexagonal close-packed (h.c.p.) structure of an iron crystal is formed by stacking layers of iron atoms in alternate layers (ABABA...) so that atoms in the second layer (B) sit in the holes formed by the atoms in the first layer (A). In the minimum-volume cell required to fill all the space with an h.c.p. array of atoms, axis *a* equals axis *b* in length, the *c* axis is normal to the *a*-*b* plane, and, for ideal h.c.p.,  $c = 1.633a$ . Steinle-Neumann *et al.*<sup>4</sup> find that seismic waves travel 12% faster in the *a*-*b* plane than along the *c* axis of an h.c.p. crystal. This result can explain the observed difference in seismic wave speeds through the inner core if there is some preferential alignment of h.c.p. iron grains. Buffett and Wenk<sup>5</sup> simulate the electromagnetic stresses acting on the inner core that cause partial alignment of h.c.p. iron grains (texture) in a way that explains the difference in wave speeds.



predict that its crystal structure will distort unexpectedly when subjected to the conditions at the centre of the Earth. Together with a model of the orientation of iron crystals in the inner core by Buffett and Wenk<sup>5</sup> (see page 60), these results can explain the observed seismic anisotropy.

The inner core is believed to be made up of iron crystals in the hexagonal close-packed (h.c.p.) structure (Fig. 1b). So the observed seismic anisotropy must arise from differences in the elastic response of h.c.p. iron — that is, how fast compressional waves travel in different directions through the crystal. (Acoustic anisotropy is the rule for all crystal structures, rather than the exception.) To understand the microstructural origin of this elastic anisotropy requires knowledge of the properties of iron at extreme pressures (about 330 gigapascals) and temperatures (about 6,000 K). One problem dogging progress so far has been that these pressure and temperature conditions are hard to create simultaneously, whether in the laboratory or in theoretical calculations. An earlier attempt at computer simulation of elasticity from first principles, for example, had to ignore the effects of temperature altogether, settling for absolute zero<sup>6</sup>.

So why do the new calculations work? Steinle-Neumann *et al.*<sup>4</sup> have produced a detailed theoretical prediction of the elastic properties of pure iron at core pressure and temperature that uses a quantum-mechanical treatment of the electrons in iron interacting both with the slower-moving atomic nuclei and with each other. This method, known as density-functional theory (DFT), has a long pedigree and has underpinned many accurate predictions of crystal structure and elasticity at low temperatures. But to calculate high-temperature properties, the thermal vibrations of the atomic nuclei must also be considered. The frequencies of these vibrations have a range that depends in turn on the quantum-mechanical forces exerted by the electrons. So a correct treatment of the nuclear motion is far more demanding than for nuclei assumed to be stationary at absolute zero.

Steinle-Neumann *et al.* solve this problem by resurrecting a technique known as the particle-in-a-cell method (used well over 30 years ago when atomic modelling was limited by the available computational power) and combining it with much more sophisticated electronic-structure methods. They have been able to include the effects of thermal vibration, and predict the effects of temperature on iron up to 8,000 K. Their work complements recent calculations on the high-temperature thermodynamics of iron at core conditions<sup>7</sup>.

The authors find that the high temperatures significantly increase the ratio of the lengths of the unit cell, known as  $c/a$ , of a single crystal of h.c.p. iron (Fig. 1b). This

change in  $c/a$  produces changes in the elasticity that in turn control the speed of seismic waves for different propagation directions in the crystal. At core temperatures of 6,000 K, compressional waves travel about 12% faster in the basal ( $a$ - $b$ ) plane of h.c.p. iron than along the  $c$ -axis (Fig. 1b). This elastic anisotropy is much stronger than expected and means that about 30% of iron crystals in the core need to be preferentially aligned, easing requirements suggested by previous calculations for almost 100% alignment. The calculation nicely explains the high Poisson's ratio — the ratio of material strain perpendicular and parallel to an extensional force — of the inner core.

But calculating the properties of a single iron crystal is not enough to determine the elastic properties of the whole inner core. It also needs a model describing the collective evolution of the orientations (texture) of many single crystals (grains) in the core. Texture can be generated by forces acting on and restructuring the core material, or it can be intrinsic to the crystal growth process. No mechanism has yet been proposed that satisfies all geodynamic criteria.

Buffett and Wenk<sup>5</sup> offer a new mechanism: electromagnetic forces acting on the inner core that arise from generation of the Earth's magnetic field in the outer core. The authors generate a model of the flow response of the inner core under magnetic stress and calculate the induced plastic (irreversible) deformation. Their model includes a step-by-step statistical simulation of grain growth from an initial random orientation, and takes into account mechanisms by which planes of atoms slide over one another, and the potential recrystallization of iron at high temperature. Calculations of the average seismic wave speeds using this texture model and the updated elastic constants of Steinle-Neumann *et al.* produce a travel-time difference in agreement with seismic observations.

As Buffett and Wenk point out, many other geodynamic effects have yet to be incorporated into models of inner-core growth, and several physical parameters used in their model are not reliably known. Moreover, the observed variation of the seismic anisotropy with depth is hard to reproduce and reflects the difficulty of determining the rate of strain accumulation relative to inner-core growth. Including the effects of high-temperature recrystallization in the model also seems to lead to stronger anisotropy than is observed.

Nonetheless, high-pressure experiments have yet to catch up and provide comparable elasticity data for iron at core temperatures. Such experiments might address the assumption that a small deviation from an inner core of pure iron has a negligible effect on elasticity and grain growth. Coupled with the prospects for clearer three-dimensional images from seismology, it may then only be a matter of time before another round



#### 100 YEARS AGO

"I returned, and saw under the sun, that the race is not to the swift, nor the battle to the strong," wrote the wise man. Writing in the same prophetic vein, M. J. de Bloch in the current *Contemporary Review*, and Mr. H. G. Wells in the *Fortnightly* for September, depict in graphic colours the transformation which the immediate future will witness in the methods of warfare. Both writers are convinced that the military tactics of the past are irretrievably dead. The effective soldier of the future will be a man whose capacity for individual action has been cultivated and developed... Mr. Wells takes into account the resources which modern science has made available for the business of war, and proceeds to anticipate the most likely directions that future advances will take. Of one thing he leaves his reader in no doubt, victory is bound to be with the nation that most sedulously attends to the education of its people in the scientific method. The great war of the future will be fought by citizens familiar with destructive instruments of precision, who have learnt to utilise all the accessory helps which science is gradually perfecting.

From *Nature* 5 September 1901.

#### 50 YEARS AGO

According to Dr. Grantly Dick Read, the influence of fear on social structure is not well understood or appreciated... Practically the whole basis of human communal existence today, he suggests, is influenced by fear; man's mental and physical health vary as to his ability to remain balanced in the face of fear. Dr. Dick Read also suggests that children are not only emotionally influenced from the moment of their birth but even from the moment of conception. At fifteen weeks the reflexes of the foetus indicate individuality of behaviour. From birth the child is regularly subjected to stimuli which induce fear and lead to insecurity. According to its nature the child risks rebellion in violence or excessively anti-social behaviour, or resolves into that submissive state when even attempted activity of any sort, or any expression of its individuality, is paralysed by the deeply implanted fear of consequences. Dr. Dick Read believes that much of this unnecessary fear with its harmful effects on the individual and society would be eliminated if the fear of child-birth in women were eliminated.

From *Nature* 8 September 1951.

of revisionism, but at least the fundamentals appear to be in place.

Andrew Jephcoat and Keith Refson are in the Department of Earth Sciences, University of Oxford, Oxford OX1 3PR, UK.  
e-mails: andrew@earth.ox.ac.uk  
keith@earth.ox.ac.uk

1. Creager, K. C. *J. Geophys. Res.* **104**, 23127–23139 (1999).
2. Vidale, J. E. & Earle, P. S. *Nature* **404**, 273–275 (2000).
3. Tromp, J. *Annu. Rev. Earth Planet. Sci.* **29**, 47–69 (2001).
4. Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Güleren, O. *Nature* **413**, 57–60 (2001).
5. Buffett, B. A. & Wenk, H.-R. *Nature* **413**, 60–63 (2001).
6. Stixrude, L. & Cohen, R. E. *Science* **267**, 1972–1975 (1995).
7. Alfé, D., Price, G. D. & Gillan, M. J. *Phys. Rev. B* **64**, 45123–45139 (2001).

## Cytoskeleton

# Evolution in bacteria

Harold P. Erickson

Actin is a major component of the cytoskeleton in yeast, plant and animal cells, but when did it evolve? The discovery of a bacterial protein that forms actin-like filaments suggests an answer.

All eukaryotic cells, from yeast to plants to animals, have an internal framework called the cytoskeleton. The three types of strut that make up this framework — microtubules, actin filaments and intermediate filaments — not only provide cells with mechanical support, but also serve as tracks for motor molecules to move along. Bacteria were long thought to lack cytoskeletal filaments, suggesting that the cytoskeleton might have evolved after the first primitive eukaryotic cell developed from its bacterial origins.

Over the past few years, however, microtubules have been traced back to the bacterial protein FtsZ, which forms filaments that are involved in cell division. The origin of actin has been more obscure (and even less is known about the evolution of intermediate filaments). Now, work by van den Ent and colleagues<sup>1</sup> (on page 39 of this issue), along with a paper by Jones and co-workers<sup>2</sup>, shows that the protein MreB forms an actin-like cytoskeleton in bacteria, providing a strong argument that the actin cytoskeleton had largely evolved before eukaryotes developed.

Several bacterial proteins are related to actin, as judged by similarities in their amino-acid sequences. But although this 'homology' shows that the proteins evolved from a common ancestor, it says nothing about their functions. In fact, these bacterial relatives have widely different functions, most of which are unrelated to the cytoskeleton. Hexokinase is an enzyme; DnaK is a 'chaperone' protein that helps other proteins to fold correctly; and FtsA functions in cell division, although perhaps not as a cytoskeletal protein. MreB, another bacterial relative of actin, is somehow involved in controlling the shape of *Escherichia coli* cells<sup>3</sup>, but its precise function has been uncertain. If one of these actin homologues is a direct evolutionary precursor of actin, it might be expected to assemble into filaments. The new studies<sup>1,2</sup> show that MreB does just that.

Both groups<sup>1,2</sup> noted that, in terms of sequence, MreB is closer to actin than are the other bacterial homologues. Jones *et al.*<sup>2</sup>

used light microscopy to examine MreB and a closely related protein, Mbl, in *Bacillus subtilis* cells. They found that these proteins form separate helical bands that spiral around inside the cell, just under the membrane. They went on to show that depletion of either protein causes defects in cell shape: MreB affects mainly the width, and Mbl the length, of the rod-shaped *B. subtilis*. Similarly, deletion of MreB in *E. coli* cells results in a marked switch from a rod shape to a sphere<sup>3</sup>.

To generalize these findings, Jones *et al.* scanned the numerous bacterial and archaeal genomes that have been sequenced. (Bacteria and archaea are the two groups of prokaryotes — organisms without membrane-bound nuclei.) They found that prokaryotes with a rod-like, filamentous or helical shape generally have one or more *mreB* genes, but spherical bacteria (cocci) have none. So the default shape for a bacterium seems to be a sphere; other types of shape are produced by an MreB-based cytoskeleton.

The spiral bands seen by light microscopy were suggestive of a cytoskeleton with a filamentous substructure, but this needed to be demonstrated. van den Ent *et al.*<sup>1</sup> do exactly this, and more. They first show that the *Thermotoga maritima* MreB protein forms polymers *in vitro*. Electron microscopy revealed that the polymers are thin 'protofilaments' (consisting of a single row of MreB proteins assembled end-to-end) that line up side-by-side to form pairs or larger parallel arrays. The spacing between MreB proteins along the protofilaments is 51 Å, close to the 55-Å spacing between subunits in eukaryotic actin.

But there is one striking difference. In eukaryotes, two actin protofilaments coil around each other to form a helical actin filament, whereas the MreB protofilaments are straight. Although biologists interested in the cytoskeleton have always thought of the helical structure as fundamental to actin, van den Ent *et al.*<sup>1</sup> present an intriguing argument that the natural shape of an isolated actin protofilament may in fact be straight, but that the two protofilaments in an actin

filament are distorted into helices by the lateral contacts that hold them together.

The authors<sup>1</sup> went on to crystallize MreB and determine its atomic structure. The structure confirms the similarity of MreB to actin, and these particular crystals also provide an amazing bonus: the MreB subunits are already assembled into protofilaments. The crystals thus reveal the structure, at atomic resolution, of the interface between subunits in the protofilaments. This interface has not been seen in crystals of actin, and has been deduced only by modelling. So the MreB structure provides the first detailed look at the interface that underlies actin-like filaments.

The helical bands seen by light microscopy<sup>2</sup> probably represent a network or a parallel bundle of MreB protofilaments. But how many protofilaments are there across the width of each band? Jones *et al.*<sup>2</sup> show that there are 8,000 molecules of MreB and 13,000 of Mbl in an average bacterium. Assuming that both of these proteins form protofilaments with the 51-Å spacing observed by van den Ent *et al.*<sup>1</sup>, this is enough to assemble 41 µm and 66 µm of MreB and Mbl protofilaments, respectively. The bacteria are about 1 µm in diameter, giving a circumference of 3 µm. So, as the helical bands are judged to make one or two turns, they would have a length of 3–6 µm. What this means is that the bands seen by light microscopy might be about 10 protofilaments wide, which could provide significant mechanical rigidity.

It seems likely that the MreB-based cytoskeleton determines cell shape by generating a force against the membrane. How might it do this? In eukaryotes, a molecular motor known as myosin, which moves along actin tracks, can generate contractile force. But no bacterial motor proteins have been identified, so a similar mechanism seems unlikely to operate in bacteria. A force could instead be generated simply by the filaments' mechanical rigidity, which would prevent them from bending. A third mechanism of force generation in eukaryotes is based on filament assembly: as new subunits are added to the ends of filaments, they push against the membrane. This is the basis for protrusion of the leading edges of amoebae and amoeboid animal cells as they move across a substrate<sup>4</sup>. Perhaps this fundamental process of cell motility, as well as the actin cytoskeleton itself, evolved in bacteria before eukaryotes formed.

Harold P. Erickson is in the Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA.

e-mail: h.erickson@cellbio.duke.edu

1. van den Ent, F., Amos, L. A. & Löwe, J. *Nature* **413**, 39–44 (2001).
2. Jones, L. J., Carballido-Lopez, R. & Errington, J. *Cell* **104**, 913–922 (2001).
3. Doi, M. *et al.* *J. Bacteriol.* **170**, 4619–4624 (1988).
4. Pollard, T. D., Blanchoin, L. & Mullins, R. D. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 545–576 (2000).



## Supramolecular chemistry

# Going for gold

Hubert Schmidbaur

Molecules containing gold atoms are not expected to form metallic bonds. But there is growing evidence that there are interactions between the gold atoms that are similar in strength to hydrogen bonds.

When gold forms a bond with another atom, it simply has to donate one electron to attain an extremely stable electronic state. Conventional wisdom says that atoms in such a state should not engage in any further external bonding. But accumulating experimental evidence<sup>1</sup> and state-of-the-art quantum-chemical calculations<sup>2</sup> have shown that there are surprisingly strong metal–metal interactions between gold atoms in molecular complexes. These interactions appear to be similar in strength to hydrogen bonds, and chemists have started to use them to design new structures with unusual physical properties. For example, Bachman and colleagues<sup>3</sup> report in the *Journal of the American Chemical Society* the construction of striking ‘rotator phases’ of gold complexes based solely on their metal–metal interactions. The behaviour of gold complexes in these rotator phases could potentially be modified to produce interesting and useful structures, such as metal-based liquid crystals or photoluminescent materials.

Most gold compounds have a linear, rod-like structure. As more and more crystal structures of complexes containing such rod-like structural elements became known in the

1980s and 1990s, chemists were surprised by the tight aggregation of the components, which always appeared to minimize the distances between the gold atoms. It seemed gold was literally always drawn to gold, and the term ‘aurophilicity’ was coined to describe this new phenomenon in structural chemistry<sup>4,5</sup>.

To form chemical bonds, atoms have to donate, acquire or share electrons. As a pure element, each gold atom has one electron ( $6s^1$ ) in excess of a closed-shell configuration ( $5d^{10}$ ); this electron is shared with the neighbouring gold atoms in bulk gold metal. When this electron is donated fully to a partner to form a compound in an oxidation reaction, gold remains in a closed-shell state,  $5d^{10}$  (in what is called the +1 oxidation state), so it is no longer expected to undergo metal–metal interactions. Yet there is overwhelming evidence that aurophilic bonding often governs the organization of gold compounds. It also plays an important role in technological processes. For example,  $[\text{Au}(\text{CN})_2]^-$  is a rod-like, five-atom anion that aggregates tightly when it is adsorbed on the surface of activated carbon, making the extraction of gold cyanide from leaching brines more efficient<sup>1</sup>. Quantum-chemical

calculations, which consider relativistic and correlation effects, have been instrumental in estimating the preferred geometry of these aggregates and their relative energies<sup>2</sup>.

Weak interactions, such as aurophilic and hydrogen bonding, cannot dictate molecular structures in the way that strong covalent bonds do, but they have a say in the arrangement of groups of molecules — the supramolecular structure. For years, chemists worked to control molecular structure but left the supramolecular structure to chance. Only in the past few decades have they begun to tackle the difficult problem of designing materials with interesting supramolecular structures. This is where knowledge of hydrogen bonding and — more recently — aurophilic bonding in gold chemistry comes in<sup>6–8</sup>.

In their paper, Bachman *et al.*<sup>3</sup> study *n*-alkylisocyanide complexes of gold chloride of the type  $(\text{R-NC})\text{AuCl}$ , (where  $\text{R} = \text{C}_n\text{H}_{2n+1}$  and  $n = 1–11$ ). The linear array of the atoms  $\text{Cl-Au-C-N}$  in the metal complex makes the molecules behave like flexible hydrocarbon chains with a rod-like end-group containing the aurophilic gold atom. When crystallized, the arrangement of the molecules follows a pattern that brings the gold atoms of neighbouring molecules close together (about 3.5 Å apart) with adjacent molecules aligned in opposite directions (Fig. 1). These zigzag chains pack in what is referred to as a herring-bone structure, as seen for other long-chain hydrocarbons bearing functional groups. Similar structures with a well-defined bilayer motif are made by unbranched alcohols,  $\text{C}_n\text{H}_{2n+1}\text{OH}$ , but in this case the chains are attached to each other through hydrogen bonds between the hydroxyl (OH) groups<sup>9</sup>.

The structural analogy between the low-temperature phases of the alcohols and the (isocyanide)gold chloride complexes is further proof of the concept that hydrogen bonding and aurophilic bonding are similar in their binding energies and directionality<sup>6</sup>. At the supramolecular level, the structures formed by chains of unbranched alcohols can develop ‘rotator phases’. Rotator phases are intermediate between the ordered crystal and the isotropic (disordered) melt. In this phase, the molecules have additional freedom of ‘rotator’ motion, which may be thought of as the chains extending — untwisting like an elastic spring — as the molecular lattice becomes less tight and rigid. These peculiar structural features generally lead to anomalously large thermal expansion, isothermal compressibility and heat capacity.

The presence of rotator phases in alcohols prompted Bachman *et al.* to investigate whether  $(\text{R-NC})\text{AuCl}$  molecules develop similar rotator phases<sup>10</sup>. They found that above a temperature of around 50 °C the crystalline (isocyanide)gold chloride complexes have physical properties characteristic of

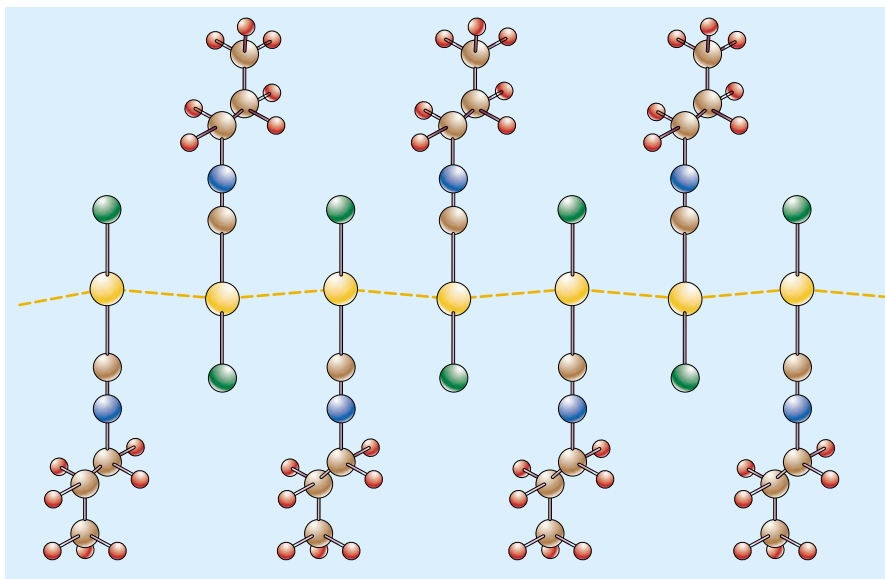


Figure 1 Supramolecular structure formed by bonding between gold atoms. Bachman *et al.*<sup>3</sup> demonstrate that aurophilic bonding between the gold atoms in linear gold chloride complexes ( $\text{C}_3\text{H}_7\text{NCAuCl}$ ) makes them align in opposite directions in a long zigzag chain. (Carbon atoms are shown in brown, hydrogen atoms are red, nitrogen atoms are blue, gold atoms are yellow, and chloride is green.) Several of these zigzag chains stack together to create a bilayer structure similar to the bilayers formed by hydrogen-bonded chains of alcohols. With longer alkyl chains, ‘rotator’ phases are formed above a transition temperature.

rotator phases. This is the first observation of such phases induced by any type of direct metal–metal bonding. Above the transition temperature the materials have been shown by polarized microscopy to be crystalline (with the individual molecules having cross-sectional areas typical of alkyl chains) but gentle mechanical pressure can deform them.

The new gold-based rotator phases are unique in that mesogenic (liquid-crystal) properties are induced in the absence of traditional mesogenic units such as aromatic rings. It remains to be seen whether these properties will lead to useful applications, perhaps in the areas of liquid-crystal and sensor technology. ■

Hubert Schmidbaur is at the *Anorganisch-chemisches Institut, Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany.*

*e-mail: h.schmidbaur@lrz.tum.de*

1. Schmidbaur, H. (ed.) *Gold: Progress in Chemistry, Biochemistry and Technology* (Wiley, Chichester, 1999).
2. Pykkö, P. *Chem. Rev.* **97**, 579–636 (1997).
3. Bachman, R. E., Fioritto, M. S., Fetics, S. K. & Cocker, T. M. *J. Am. Chem. Soc.* **123**, 5376–5377 (2001).
4. Schmidbaur, H. *Chem. Soc. Rev.* **24**, 391–401 (1995).
5. Schmidbaur, H. *Gold. Bull.* **23**, 11–21 (1990).
6. Schmidbaur, H. *Gold. Bull.* **33**, 3–9 (2000).
7. Braga, D., Grepioni, F. & Desiraju, G. R. *Chem. Rev.* **98**, 1375–1390 (1998).
8. Schneider, W., Bauer, A. & Schmidbaur, H. *Organometallics* **15**, 5445–5447 (1996).
9. Wang, J.-L. et al. *J. Am. Chem. Soc.* **116**, 1192–1197 (1994).
10. Sirota, E. B. & Wu, X. Z. *J. Chem. Phys.* **105**, 7763–7773 (1996).

## Archaeology

# Out in the cold

John A. J. Gowlett

Humans are very adaptable: during the last ice age, they apparently lived within the Arctic Circle. The discovery suggests that, although cold, the region was probably not covered in ice at the time.

Archaeological finds described by Pavlov and colleagues on page 64 of this issue<sup>1</sup> show for the first time that humans were present north of the Arctic Circle almost 40,000 years ago, in the last ice age. The idea of people living in a land gripped by an ice age goes back to nineteenth-century France, but the new finds extend both the geographic and the temporal range of the phenomenon. The results should also rekindle debate about

the effects of the climate on the movements of early human populations.

Pavlov et al.<sup>1</sup> carried out fieldwork at a site in the Russian Arctic known as Mamontovaya Kurya, which dates to Middle to Upper Palaeolithic times, some 35,000–40,000 years ago. Their finds comprise various stone tools and over a hundred mammalian bones, as well as a mammoth tusk bearing cut marks that were apparently made by tools. The age of the tusk was determined by a radiocarbon-dating technique known as accelerator mass spectrometry, illustrating the power of this technique for dating artefacts directly rather than by the age of the sediments in which they are found.

It is not possible from these finds to determine whether they were left by Neanderthals or by some of the first modern humans in Europe, but this is equally true of most contemporary artefacts further south. In the broader scheme of things, knowing who made the tools is less important than simply knowing that someone was adapted to the cold conditions. This is significant because all evidence from recent foragers (such as Inuit or Siberian Yukaghir) suggests that adaptation to northern climes requires high levels of technological and social organization.

That said, it would be interesting to know whether these people were Neanderthals or early 'anatomically modern' humans. If they were Neanderthals, this provides further support — along with their anatomical adaptations and the height and remoteness of many of the sites at which Neanderthal artefacts have been found — of the Neanderthals' rugged durability and extensive capabilities<sup>2</sup>. Their high degree of meat-eating,

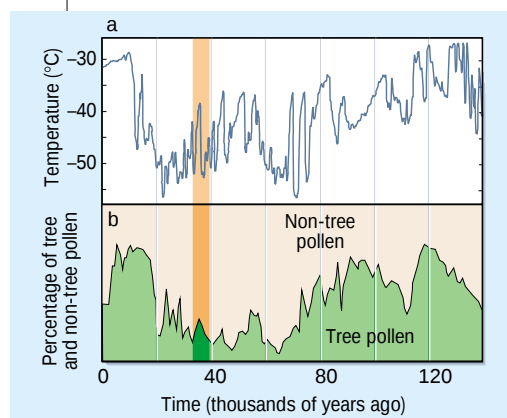
indicated by recent studies of stable carbon and nitrogen isotopes in bone<sup>3</sup>, also suggests a specialized socioeconomic adaptation, perhaps developed over a long period in environments rich in animals but limited in plant resources.

If, on the other hand, these ice-age people were modern humans, then this is evidence of a remarkably rapid advance to the north — modern humans had only just set foot in the southeast of Europe. Pavlov et al.<sup>1</sup> incline towards this view — the nearest archaeological finds to the south, along the Pechora River near the Urals<sup>4,5</sup>, are allied more closely with those Upper Palaeolithic traditions associated with modern humans than with the Middle Palaeolithic toolkits more commonly associated with Neanderthals.

The bones and artefacts found at Mamontovaya Kurya suggest that the north-east must have been relatively dry and ice-free in this period of the ice age. These finds are one outcome of a major interdisciplinary study that has also shown that, for most of the time, the ice sheets of the last glaciation were far more restricted on their eastern flank than is sometimes suggested. Support for the existence of large ice-free areas also comes from Finland<sup>6</sup>, where direct radiocarbon dating of key evidence — this time a series of mammoth teeth — establishes the presence of large animals between 22,000 and about 40,000 years ago. The existence of large animals implies that the environment was steppe-like, consisting of open grassland.

Icy cold, however, it was. Temperature estimates derived from variations in the <sup>18</sup>O content of Greenland ice cores such as GRIP2 (at a latitude of about 72° north) show wild fluctuations throughout the middle of the last glaciation<sup>7</sup> (Fig. 1), but it was always at least 10 °C colder than today. The dominant feature of the time from 60,000 to 30,000 years ago was a series of saw-toothed temperature fluctuations of up to 15 °C. Similar temperatures were found in intensive research across north-central Europe, using indicators such as beetle remains and pollen as proxy evidence<sup>8</sup>. These indicate an average annual temperature of –1 °C in the Netherlands from 50,000 to 41,000 years ago, with the coldest month being at least 10 °C below this<sup>9</sup>.

Yet, despite this newly detailed backdrop, recent archaeology-based discussion about the Neanderthals has not — with certain exceptions<sup>10</sup> — been concerned primarily with climate. The emphasis has been on chronology, population movements, and the nature of cultural contact (if indeed there was any) between Neanderthals and incoming modern humans (Box 1, overleaf). There has been good reason to focus on the cultural changes that occurred in the past 40,000 years, as this time period includes the more rapid developments of the Upper Palaeo-



**Figure 1 Climate and early human populations.** Pavlov et al.'s finds from Mamontovaya Kurya<sup>1</sup> are dated to the period shown by the orange shading. This is set against the backdrop of: a, the temperature record of the last ice age, obtained from analysis of Greenland ice cores such as GRIP2 (ref. 7); and b, the pollen record from Lake Kopais, Greece<sup>13</sup>. These and other findings suggest that the ice sheets of the last glaciation were, most of the time, fairly restricted on the northeastern flank, allowing people (Neanderthals or early modern humans) to settle there. (The timescale is only approximate for b.)



## Box 1 Neanderthal prehistory

Neanderthals are the best known group of early hominids; they lived in Europe and parts of Asia, and survived until about 30,000 years ago. Hominid specimens from about 300,000 years ago already had some characteristics of Neanderthals, but it is not known how much further back their roots go. Some early humans entered Europe at least 800,000 years ago, but later technological innovations shared with other regions suggest that new populations may have entered the continent.

Better known are the later, 'classic Neanderthal' specimens from the last glaciation. The features of these Neanderthals, who were around from about 100,000 to 30,000 years ago, are more uniform than those of earlier populations, suggesting that there may have been a

population crash (with a subsequent 'bottleneck') around the start of this period<sup>15</sup>. All Neanderthals were strongly built, with long, low crania and large faces. Damage to their skeletons attests to the fact that their lives were hard. If they were as carnivorous as seems likely, they would have had many encounters with dangerous prey.

After 150 years of debate, and despite a steady flow of new knowledge, the Neanderthals' position alongside anatomically modern humans remains uncertain<sup>16</sup>. The means by which Neanderthals were eventually replaced by modern humans is fiercely contested, as is the degree of genetic separation. Many specialists see Neanderthals as a distinct species, contributing little to either the gene pool or the culture of later populations.

But modern human and Neanderthal remains, genetics and tool traditions all show intriguing continuities<sup>17</sup>.

Less controversial is the new willingness to admit that Neanderthals had qualities that showed their 'humanity'. For example, Neanderthals are contenders for the first display of caring behaviour: a crippled individual at Shanidar in Iraq was clearly sustained by the support of others. But it is not easy to pick out behavioural patterns that were distinctive of Neanderthals — burials, bone-based tools and symbolism are all found earlier in populations of anatomically modern humans. The continued debate over the relationship between Neanderthals and modern humans is stimulating, but should not mask enormous advances in dating, genetics and other forms of analysis. **J.A.J.G.**

lithic era and falls within the range of radio-carbon dating. Climate change has seemed less important because different Neanderthal populations successfully made distinct adaptations to different regions<sup>11</sup>, and these adaptations remained roughly the same for 200,000 years<sup>12</sup>.

But climate may have its moment again. The dramatically spiky record from ice cores in the interval from 60,000 to 40,000 years ago, together with pollen evidence, implies that steppe environments moved up and down rapidly from southeast Europe to the far north, and suggests that climate change could have been crucial in promoting population movement and cultural change. In 'warmer' parts of the ice age, as Pavlov *et al.*<sup>1</sup> show, fauna-rich steppe environments and humans apparently reached the Arctic. During colder intervals, wooded environments gave way to steppe even in Greece<sup>13</sup>. In the Last Glacial Maximum, 20,000 years ago, conditions were so ferociously cold that even modern humans were driven down towards the south of France<sup>14</sup>. Indirectly, such responses may help to explain the southward expansion of Neanderthals into the Middle East around 60,000 years ago, and (perhaps) the similar spread of Upper Palaeolithic Aurignacian human populations around 30,000 years ago.

The new finds<sup>1</sup> show that humans had a hold on the north, if only for a short time. Although there are questions to be answered,

the artefacts illustrate both the capacity of early humans to do the unexpected, and the value of archaeologists researching in unlikely areas. ■

John A. J. Gowlett is in the Department of Archaeology, University of Liverpool, Liverpool L69 3BX, UK.

e-mail: [gowlett@liverpool.ac.uk](mailto:gowlett@liverpool.ac.uk)

- Pavlov, P., Svendsen, J. I. & Indrelid, S. *Nature* **413**, 64–67 (2001).
- Hayden, B. J. *Hum. Evol.* **24**, 113–146 (1993).
- Bocherens, H. *et al. J. Archaeol. Sci.* **26**, 599–607 (1999).
- Ivanova, I. K. *L'Anthropologie* **73**, 5–48 (1972).
- Allsworth-Jones, P. in *The Emergence of Modern Humans* (ed. Mellars, P.) 160–242 (Edinburgh Univ. Press, 1990).
- Ukkonen, P., Lunkka, J. P., Jungner, H. & Donner, J. *J. Quat. Sci.* **14**, 711–714 (1999).
- Johnsen, S. J. *et al. J. Quat. Sci.* **16**, 299–307 (2001).
- Vandenberghe, J., Kasse, K. & Coope, R. (eds) *J. Quat. Sci.* **13**, 361–497 (1998).
- Huijzer, B. & Vandenberghe, J. *J. Quat. Sci.* **13**, 391–417 (1998).
- Bar-Yosef, O. in *Paleoclimate and Evolution, with Emphasis on Human Origins* (eds Vrba, E. S., Denton, G. H., Partridge, T. C. & Burckle, L. H.) 507–523 (Yale Univ. Press, New Haven, CT, 1995).
- Gamble, C. *The Palaeolithic Societies of Europe* (Cambridge Univ. Press, 1999).
- Mellars, P. *et al. Curr. Anthropol.* **40**, 341–364 (1999).
- Okuda, M., Ysuda, Y. & Setoguchi, T. *Boreas* **30**, 73–82 (2001).
- Bocquet-Appel, J.-P. & Demars, P.-Y. *J. Archaeol. Sci.* **27**, 551–570 (2000).
- Bilborough, A. in *The Hominids and their Environment during the Lower and Middle Pleistocene of Eurasia: Proceedings of the International Conference of Human Palaeontology, Orce, 1995* (eds Gibert, J., Sanchez, F., Gibert, L. & Ribot, F.) 311–315 (Museo de Prehistoria y Paleontología, Orce, 1999).
- Fox, R. G. (ed.) *Curr. Anthropol.* **39** (suppl.), S1–S189 (1998).
- Churchill, S. E. & Smith, F. H. *Yb. Phys. Anthropol.* **43**, 61–115 (2000).

## Daedalus

### Pay for the Internet

The collapse of many Internet dotcom companies reflects the deep unease of many consumers at trusting their credit-card numbers to the insecurity of the Internet. How, then, to bring alive the Internet's commercial potential? Micro-payments have somehow not taken off. Daedalus reckons that a consumer will trust to the Internet what he might lose on a bet — maybe US\$50 or so. Rogue sites that charge more than this, in one or many payments, must be discouraged.

DREADCO Financial Services is therefore inventing a special low-limit card or cheque for Internet use. Their small limit cannot be overexploited, so the risk is low. It may feature an exact cut-off negotiated automatically with the company beforehand.

Thus the great dream — that of our own personal library from the Internet — would come true at last. The big publications, like *Nature* and *The Times*, will continue as before, for those who need them. But people who visit their websites, and those of all other publications too, will note those articles that seem interesting, and amass them for small proportional payments, controlled by the DREADCO card. A software 'gopher' could do this or haggle over it all the time. It would store its haul on a hard disk, a vast improvement on a stack of old papers. Occasionally you might glance at the screen to see what you had caught, and you could print out your results. Those for whom citations are magical, in support of personal glory, would relish the chance to amass them. The whole media scene would be one vast newspaper or magazine, with perfect indexing, no wasted time, and maybe the deletion of items no longer of interest.

Why doesn't this happen now? The academic media are deeply divided about how freely to release their papers onto the Internet, or what time delay to adopt before doing so. But once most publishers had taken the plunge, the worth of any delay would soon be shown by market forces, clear to everyone. Daily newspapers would lose value in a few days, and weeklies in a few weeks; academic journals of more lasting value should hold that value longer. Daedalus reckons that when a journal has amassed more citations than it quotes itself, its value starts to decay. But journal publication would remain useful, partly because editorial scrutiny weeds out the nonsense, and partly because anyone specializing in the area is better off subscribing to the journal.

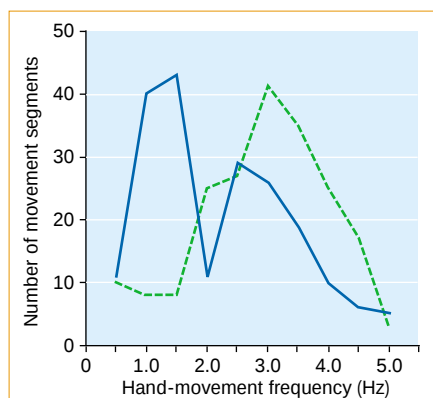
David Jones

# Language rhythms in baby hand movements

Hearing babies born to deaf parents babble silently with their hands.

The vocal babbling sounds universally uttered by healthy babies at around 7 months of age are fascinating, and have been interpreted as reflecting both the origins of language production in humans<sup>1</sup> and the vestiges of the evolutionary origins of language in our species<sup>2</sup>. Here we study the hand movements of hearing babies born to profoundly deaf parents and find that these children produce a class of hand activity that is distinct from other uses of their hands and which contains the specific rhythmic patterns of natural language ('silent' babbling). Our findings support the idea that babies are sensitive to rhythmic language patterns and that this sensitivity is key to launching the process of language acquisition.

The biological basis of babbling has been debated for decades. One possibility is that babbling, as in modern accounts of the origins of human language, is a purely non-linguistic motor activity that results from the opening and closing of the mouth and jaw<sup>3–6</sup>. Alternatively, babbling could be a linguistic activity that reflects babies' sensitivity to specific patterns at the heart of human language and their capacity to use them<sup>7–9</sup> — particularly the rhythmic patterns that bind syllables, the elementary units of language, into baby babbles, and then into words and sentences.



**Figure 1** Hand-movement frequencies calculated for the rhythmic hand activity of sign-exposed (full line) and speech-exposed (dashed line) babies across all ages; for each group, 400 movement segments (200 per group) were randomly selected. Only sign-exposed babies had a bimodal distribution of movement frequencies: the first mode (left peak) falls at around 1 Hz (range, 0.5–1.5 Hz) and the second mode (right peak) falls at around 2.5 Hz (range, 2.0–3.0 Hz). In contrast, hand-movement frequencies of speech-exposed babies were unimodal, falling at around 3.0 Hz (range, 2.5–3.5 Hz). Comparison of the two groups further revealed that the pattern of movement frequencies produced by sign-exposed babies was significantly different from that of speech-exposed babies at the same age (20,  $n=200$ ;  $\chi^2=389.65$ ,  $P<0.001$ ;  $\chi^2$  was calculated at 21 quarter-intervals and is shown here at half-intervals for clarity).

To test the motor and linguistic hypotheses, we studied three hearing babies who received no systematic exposure to spoken language and who instead saw only signed language from their profoundly deaf parents, and three hearing babies who were exposed to spoken language. We previously compared the capacity of hearing and deaf babies to babble in another study, in which group differences may have resulted from the babies' different sensory experiences<sup>10</sup>.

The two hearing baby groups were equal in all developmental respects, with the only difference being in the form of language input they received (by hand or mouth). Because hearing babies exposed to sign language do not use their mouth and jaw to learn speech, the motor hypothesis predicts that their hand activity should be fundamentally similar to that of hearing babies acquiring spoken language. If, however, babies are born with sensitivity to specific rhythmic patterns that are universal to all languages, even signed ones, then the linguistic hypothesis predicts that differences in the form of language input should yield differences in the hand activities of the two groups.

We recorded all babies' hand activity in three dimensions using Optotrak, an optoelectronic position-tracking system. The hand activity was carried out during presentation with objects and during game-playing in 60-min experimental sessions conducted when the babies were aged about 6, 10 and 12 months. Optotrak sensors accurately measure the trajectory and location over time of light-emitting diodes on the babies' hands with a 0.1-mm precision. Optotrak computations were carried out blind to videotape recordings of the positions of the babies' hands, which on their own are a subjective way to analyse hand movements<sup>11</sup>. Online videotapes were made of all babies independently for post-Optotrak analysis.

Optotrak analyses revealed that sign-exposed babies showed a significantly different type of low-frequency rhythmic hand activity from speech-exposed babies, as well as another type of high-frequency rhythmic hand activity that speech-exposed babies also showed and used almost exclusively (Fig. 1).

The low-frequency hand activity of sign-exposed babies was mainly generated within a tightly restricted space (Fig. 2), corresponding to the obligatory 'sign-phonetic' space in front of a signer's body that binds all linguistic expression in signed languages (82%); high-frequency hand activity was mainly outside this space (73%). Speech-exposed babies produced most of their high-frequency hand activity



**Figure 2** Learning language: a class of hand movements made by babies with profoundly deaf parents have a slower rhythm than ordinary gestures and are restricted to space in front of the body.

outside the crucial linguistic space (92%). Quantitatively, the low-frequency hand activity corresponds to the rhythmic patterning of adult sign-syllables<sup>12</sup>. We also discovered, after lifting the blind on videotape recordings, that only these low-frequency movements had the qualitative properties of silent linguistic hand babbling<sup>10</sup>.

Remarkably, and without relying on the mouth, this silent linguistic babbling was conveyed by babies' hands in a different class of movement from non-linguistic hand activity. These linguistic and motor movements are differentiated by their distinct rhythmic frequencies, which could only result if babies are able to use the specific rhythmic patterns that underlie human language.

**Laura Ann Petitto\*†, Siobhan Holowka\*, Lauren E. Sergio‡, David Ostry\*§**

\*Department of Psychology, McGill University, Montréal, Québec H3A-1B1, Canada

‡Department of Kinesiology and Health Sciences, York University, Toronto, Ontario M3J-1P3, Canada

§Haskins Laboratories, New Haven, Connecticut 06511-6695, USA

†Present address: Departments of Psychological & Brain Sciences and Education, Dartmouth College, 302 Silsby Hall, Hanover, New Hampshire 03755, USA

e-mail: laura-ann.petitto@dartmouth.edu

- Locke, J. L. *The Child's Path to Spoken Language* (Harvard Univ. Press, Cambridge, Massachusetts, 1993).
- Lieberman, P. *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought* (Harvard Univ. Press, Cambridge, Massachusetts, 2000).
- MacNeilage, P. F. & Davis, B. L. *Science* **288**, 527–531 (2000).
- Locke, J. L. *Science* **288**, 449–451 (2000).
- Studdert-Kennedy, M. G. in *Cognition and the Symbolic Processes: Applied and Ecological Perspectives* (eds Hoffman, R. R. & Palermo, D. S.) 38–58 (Erlbaum, Hillsdale, New Jersey, 1991).

6. van der Stelt, J. M. & Koopmans-van Beinum, F. J. in *Precursors of Early Speech* (eds Lindblom, B. & Zetterstrom, R.) 163–173 (Stockton, New York, 1986).
7. Pinker, S. & Bloom, P. in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (eds Barkow, J. H. et al.) 451–493 (Oxford Univ. Press, New York, 1992).
8. Jusczyk, P. W. *The Discovery of Spoken Language* (MIT Press, Cambridge, Massachusetts, 1997).

# Spectrographic imaging

## A bird's-eye view of the health of coral reefs

Almost three-quarters of the world's coral reefs are thought to be deteriorating as a consequence of environmental stress. Until now, it has been possible to evaluate reef health only by field survey, which is labour-intensive and time-consuming. Here we map live coral cover from the air by remote imaging, a technique that will enable the state of shallow reefs to be monitored swiftly and over large areas.

It is predicted that coral reefs will suffer mounting stress associated with a global increase in atmospheric carbon dioxide over the coming decades<sup>1,2</sup> and from local disturbances such as overfishing<sup>3</sup> and disease<sup>4</sup>. The most obvious effect of such stress is a decline in living coral cover, so a temporal change in cover is a good indicator of the state of health of a coral reef. However, the measurement of coral cover by field survey<sup>5</sup> is impractical on the scale of hundreds to thousands of square kilometres.

Measurements of the reflected light spectra of reef biota and substrata indicate that the dominant groups can be distinguished *in situ*<sup>6–10</sup>, but until now it has not been clear whether such spectral differences can be detected remotely from the air or from space. We acquired high-spatial-resolution (1 m × 1 m), multispectral images from the air of two reefs in the lagoon of Rangiroa Atoll, French Polynesia, by using a compact spectrographic imager. We carried out this imaging in November 1998 because coral populations had suffered significant mortality after the extreme El Niño/Southern Oscillation that occurred in the austral summer of 1997–98 (ref. 11). At the same

9. Vihman, M. M. *Phonological Development: The Origins of Language in the Child* (Blackwell, Cambridge, MA, 1996).
10. Petitto, L. A. & Marentette, P. F. *Science* **251**, 1493–1496 (1991).
11. Meier, R. P. & Willerman, R. in *Language, Gesture, and Space* (eds Emmorey, K. & Reilly, J. S.) 391–409 (Erlbaum, Mahwah, New Jersey, 1995).
12. Petitto, L. A. et al. *Proc. Natl Acad. Sci. USA* **97**, 13961–13966 (2000).

time, we carried out detailed *in situ* surveys of coral populations on each reef.

The first reef was dominated by large live and dead colonies of *Porites*, the remotely sensed spectra of which were readily distinguishable on the basis of their first spectral derivatives (rate of change of reflectance versus wavelength<sup>12</sup>) in the wavelength region 506–565 nm, as expected from *in situ* measurements<sup>8</sup>.

The habitat of the second reef was highly heterogeneous and we used the imager to estimate the percentage cover of all substrata in ten plots of 25 m<sup>2</sup> each. Although within-plot estimates of the cover of dead *Pocillopora* coral, coralline red algae and sand varied by as much as 25–29%, estimates of live coral never differed from field data by more than 10% (Table 1). At a whole-reef (interplot) scale, estimates of the mean cover of all major benthic categories differed by less than 8%, and the cover of both live and dead coral was estimated to within 3%. There were no significant differences in estimates of mean habitat cover at the 95% confidence level (paired *t*-tests).

The video surveying methods currently used by scientific divers on the Great Barrier Reef are estimated to have a 95% probability of detecting a 10% change in live coral cover from one year to the next<sup>13</sup>. Power analysis of our data indicates that 22 plots of 25 m<sup>2</sup> each would need to be surveyed by remote sensing, compared with 20 equivalent-sized plots by video camera, to achieve the same level of statistical resolution on the reefs surveyed (that is, similar sample sizes for field and remote methods). However, spectrographic images can be acquired over areas that are many times larger than those that can be surveyed underwater.

We anticipate that the application of multispectral remote sensing will signifi-

cantly improve estimates of coral cover and changes in coral cover over time. It took us 1 hour to acquire images over 92,500 m<sup>2</sup> of reef, which represents 3,700 plots of 25 m<sup>2</sup> each, compared with 3 days to survey 10 such plots underwater. Moreover, remote images are acquired as numerical data, which can be rapidly processed electronically, reducing the time needed to generate estimates of surface cover.

**Peter J. Mumby\***, **John R. M. Chisholm†**, **Chris D. Clark‡**, **John D. Hedley\***, **Jean Jaubert†§**

\*Tropical Coastal Management Studies, Department of Marine Sciences and Coastal Management, Ridley Building, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

e-mail: p.j.mumby@ncl.ac.uk

†Observatoire Océanologique Européen, Centre Scientifique de Monaco, Avenue Saint-Martin, Monaco 98000, Monaco

‡Department of Geography and Sheffield Centre for Earth Observation Science, University of Sheffield, Sheffield S10 2TN, UK

§Khaled Bin Sultan Living Oceans Foundation, 1504 East Grand River, East Lansing, Michigan 48823, USA

1. Pittock, A. B. *Am. Zool.* **39**, 10–29 (1999).
2. Kleypas, J. A. et al. *Science* **284**, 118–120 (1999).
3. Hughes, T. P. *Science* **265**, 1547–1551 (1994).
4. Harvell, C. D. et al. *Science* **285**, 1505–1510 (1999).
5. English, S., Wilkinson, C. & Baker, V. *Survey Manual for Tropical Marine Resources* (Australian Inst. Mar. Sci., Townsville, 1997).
6. Holden, H. & LeDrew, E. *Remote Sensing Environ.* **65**, 217–224 (1998).
7. Myers, M. R., Hardy, J. T., Mazel, C. H. & Dustan, P. *Coral Reefs* **18**, 179–186 (1999).
8. Clark, C. D., Mumby, P. J., Chisholm, J. R. M., Jaubert, J. & Andrefouet, S. *Int. J. Remote Sensing* **21**, 2321–2327 (2000).
9. Hochberg, E. J. & Atkinson, M. J. *Coral Reefs* **19**, 164–171 (2000).
10. Lubin, D., Li, W., Dustan, P., Mazel, C. H. & Starnes, K. *Remote Sensing Environ.* **75**, 127–137 (2001).
11. Mumby, P. J. et al. *Mar. Biol.* **139**, 183–189 (2001).
12. Tsai, F. & Philpot, W. *Remote Sensing Environ.* **66**, 41–51 (1998).
13. Carleton, J. H. & Done, T. J. *Coral Reefs* **14**, 35–46 (1995).

# Immune recognition

## A new receptor for β-glucans

The carbohydrate polymers known as β-1,3-D-glucans exert potent effects on the immune system — stimulating antitumour and antimicrobial activity, for example — by binding to receptors on macrophages and other white blood cells and activating them. Although β-glucans are known to bind to receptors, such as complement receptor 3 (ref. 1), there is evidence that another β-glucan receptor is present on macrophages. Here we identify this unknown receptor as dectin-1 (ref. 2), a finding that provides new insights into the innate immune recognition of β-glucans.

We screened a RAW264.7 complementary DNA retroviral expression library using the β-glucan-rich particle zymosan<sup>3</sup>

**Table 1 Estimates of substrate cover for live and dead reefs**

|                              | Cover (%)    |             | Mean disparity (%) | Maximum disparity (%) | Minimal detectable disparity of means (%) |
|------------------------------|--------------|-------------|--------------------|-----------------------|-------------------------------------------|
|                              | Field survey | Imagery     |                    |                       |                                           |
| Live <i>Porites</i>          | 11.5 (2.7)   | 8.8 (2.9)   | –2.7               | –9                    | 5.6                                       |
| Recently dead <i>Porites</i> | 7.8 (1.6)    | 6.8 (3.1)   | –1.0               | 18.9                  | 10.6                                      |
| Dead <i>Pocillopora</i>      | 32.5 (8.3)   | 37.2 (12.6) | 4.7                | 28.9                  | 17.3                                      |
| Red coralline algae          | 21.9 (2.4)   | 15.2 (4.8)  | 6.6                | –25.7                 | 14.3                                      |
| Sand                         | 18.9 (6.0)   | 26.8 (9.1)  | 7.9                | 27.2                  | 12.1                                      |
| <i>Halimeda</i>              | 3.9 (1.6)    | 5.1 (2.1)   | 1.2                | –9.7                  | 8.3                                       |

Field and remote estimates of substrate cover are shown for comparison (standard errors shown in parentheses). Results of pairwise *t*-test comparisons for each habitat were non-significant (*P* < 0.05). Minimal detectable difference represents the smallest disparity in mean cover between field and image estimates that would result in a significant *t*-test with 90% power. These values provide a worst-case scenario for the accuracy of remote sensing to predict mean habitat cover; actual disparities were considerably lower. Compact airborne spectrographic imager (CASI) data (10 bands) were corrected for depth variation (1–7 m) using image-derived attenuation coefficients. Substrata were predicted from unsupervised classification of spectral data and were categorized using independent field data. Each plot was identified on CASI images by triangulation to white plastic markers (4 m<sup>2</sup>) and mapped *in situ* with a resolution of 0.01 m<sup>2</sup>.



and isolated a single receptor that bound to zymosan. The DNA sequence identified the receptor as dectin-1, a small (relative molecular mass about 28,000) type-II membrane receptor with an extracellular C-type lectin-like domain fold and a cytoplasmic domain with an immunoreceptor tyrosine-based activation motif<sup>2</sup>. In contrast to its reported specificity for dendritic cells<sup>2</sup>, we found that dectin-1 was expressed in every macrophage population we examined and in more tissues than was previously reported, with the highest expression being in the liver, lung and thymus (results not shown).

By assaying the ability of different carbohydrates to block the binding of zymosan to NIH3T3 cells expressing dectin-1, we found dectin-1 to be a pattern-recognition receptor that recognizes a variety of  $\beta$ -1,3-linked and  $\beta$ -1,6-linked glucans from fungi and plants (Fig. 1a). Dectin-1 did not recognize monosaccharides (data not shown) or carbohydrates with different linkages. Laminarin and glucan phosphate, a structurally defined, immunologically active  $\beta$ -glucan<sup>4</sup>, were the most effective inhibitors; both bind to the  $\beta$ -glucan receptor on monocytes and macrophages<sup>5,6</sup>. The ability of dectin-1 transfectants to bind to zymosan was trypsin-sensitive, a well-known feature of the  $\beta$ -glucan receptor<sup>7</sup>.

The C-type lectin-like fold of dectin-1 is

similar to those of natural-killer T-cell C-type lectin domains, which lack the residues that are involved in calcium coordination and are required for carbohydrate binding in classic  $\text{Ca}^{2+}$ -dependent C-type lectins<sup>8</sup>. This is consistent with our finding that binding of dectin-1 to zymosan is independent of metal ions (results not shown).

Soluble, recombinant dectin-1 also stimulates the proliferation of T lymphocytes<sup>2</sup>. In a whole-cell binding assay, binding of T cells to NIH3T3 cells expressing dectin-1 was not inhibited by  $\beta$ -glucans (results not shown). We conclude that dectin-1 has two ligand-binding sites: one that recognizes an endogenous ligand on T cells<sup>2</sup>, and another for exogenous carbohydrates.

The  $\beta$ -glucan receptor has also been implicated in the recognition and phagocytosis of intact *Saccharomyces cerevisiae*<sup>9</sup> and of the fungal pathogen *Candida albicans*<sup>10</sup>. Both of these organisms were bound to by dectin-1 transductants in a  $\beta$ -glucan-dependent manner (Fig. 1b), consistent with the presence of  $\beta$ -1,3-linked and  $\beta$ -1,6-linked glucans within their cell walls<sup>11</sup>. Dectin-1 also mediates actin-dependent phagocytosis of zymosan, an activity that requires the cytoplasmic tail of this receptor (results not shown). Furthermore, *C. albicans* conidia were internalized (Fig. 1c), showing that dectin-1 can mediate non-opsonic phago-

cytosis of this opportunistic pathogen.

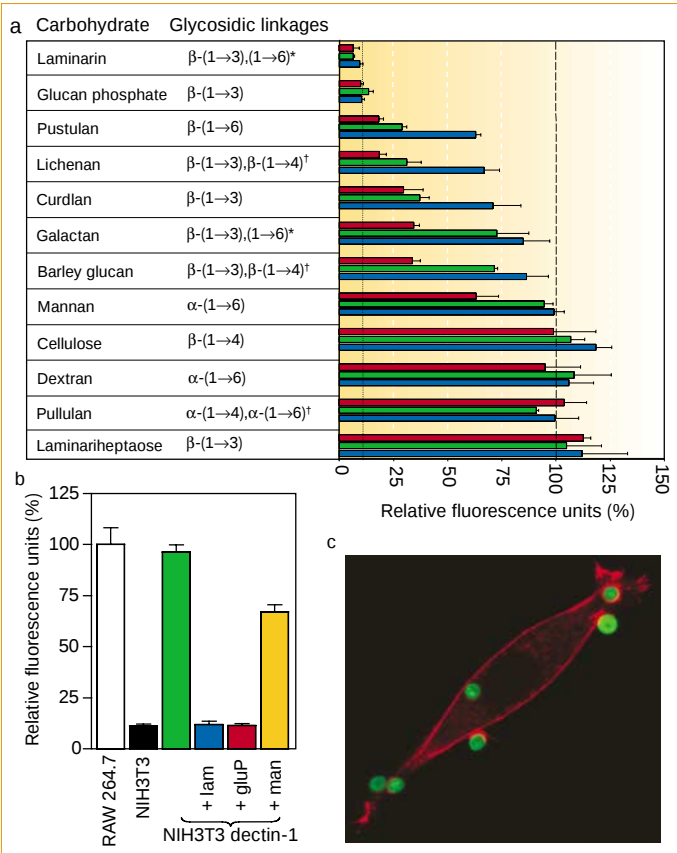
A human homologue of dectin-1 (GenBank accession number, AY009090) is similar to the murine receptor, except that it lacks an extracellular stalk region and has no sites for *N*-linked glycosylation. Binding of zymosan and *C. albicans* by the human receptor is also dependent on  $\beta$ -glucan (results not shown), indicating that it may be the functional equivalent of dectin-1. Our identification of dectin-1 as the elusive macrophage receptor for  $\beta$ -glucan resolves a long-standing mystery and will open up new opportunities to exploit the effects of  $\beta$ -glucans.

**Gordon D. Brown, Siamon Gordon**

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK  
e-mail: gbrown@molbiol.ox.ac.uk

1. Ross, G. D., Cain, J. A. & Lachmann, P. J. *J. Immunol.* **134**, 3307–3315 (1985).
2. Ariizumi, K. et al. *J. Biol. Chem.* **275**, 20157–20167 (2000).
3. Di Carlo, F. J. & Fiore, J. V. *Science* **127**, 756–757 (1958).
4. Williams, D. L. et al. *Carbohydr. Res.* **219**, 203–213 (1991).
5. Muller, A. et al. *J. Immunol.* **156**, 3418–3425 (1996).
6. Mueller, A. et al. *Glycobiology* **10**, 339–346 (2000).
7. Czop, J. K. *Pathol. Immunopathol. Res.* **5**, 286–296 (1986).
8. Weis, W. I., Taylor, M. E. & Drickamer, K. *Immunol. Rev.* **163**, 19–34 (1998).
9. Giannini, J. et al. *J. Leuk. Biol.* **54**, 564–571 (1993).
10. Janusz, M. J., Austen, K. F. & Czop, J. K. *Immunology* **65**, 181–185 (1988).
11. Bartnicki-Garcia, S. *Annu. Rev. Microbiol.* **22**, 87–108 (1968).

**Figure 1** Dectin-1 is a phagocytic pattern-recognition receptor that recognizes  $\beta$ -1,3-linked and/or  $\beta$ -1,6-linked glucans and intact yeast particles. **a**, Dectin-1-transduced cells were pre-treated with 500  $\mu\text{g ml}^{-1}$  (red bars), 100  $\mu\text{g ml}^{-1}$  (green bars) or 10  $\mu\text{g ml}^{-1}$  (blue bars) of carbohydrate before addition of fluorescently labelled zymosan particles (50 per cell). Zymosan binding was quantified by fluorometry and is expressed relative to an uninhibited control (100%; dashed line). Background binding of zymosan to untransduced NIH3T3 cells was normally about 10% (dotted line). Carbohydrates with side-chain linkages (asterisks) and/or mixed linkages (daggers) are composed of glucan polymers, except galactan (galactose monomer) and mannan (mannose monomer). **b**, Dectin-1 mediates  $\beta$ -glucan-dependent binding of fluorescently



labelled, heat-killed *Candida albicans*, which is comparable to binding by RAW264.7 macrophages. lam, laminarin; gluP, glucan phosphate; man, mannan. **c**, Immunofluorescent micrograph of dectin-1 mediating non-opsonic phagocytosis of fluorescently labelled, heat-killed *C. albicans* (green) in NIH3T3 transductants by means of actin-based phagocytic cups (red).

## Evolutionary genetics

# Clonal inheritance of avian mitochondrial DNA

We have taken a new approach to test the commonly accepted, but recently questioned, principle<sup>1,2</sup> of clonal inheritance of vertebrate mitochondrial DNA (mtDNA) by relating its inheritance to a female-specific marker of nuclear DNA. Whereas this is impossible in organisms with male heterogamy (such as mammals), we show here that genealogies of mtDNA and the female-specific W chromosome of a bird species are completely concordant. Our results indicate that inheritance of mtDNA is free of detectable recombination effects over an evolutionary timescale.

The avian W chromosome is small, has a low gene content, and is rich in heterochromatic, repetitive DNA. Most of the chromosome does not recombine<sup>3</sup>, and genes in the non-recombining part are thus exclusively and clonally transmitted by females, a situation that is confirmed by the independent evolution of paralogous genes on chromosomes Z and W<sup>4,5</sup>. Because of this clonal inheritance, we would expect the W chromosome to segregate perfectly with mtDNA, provided that the latter is also clonally transmitted only from mother to daughter, without recombination. If this is not the case,

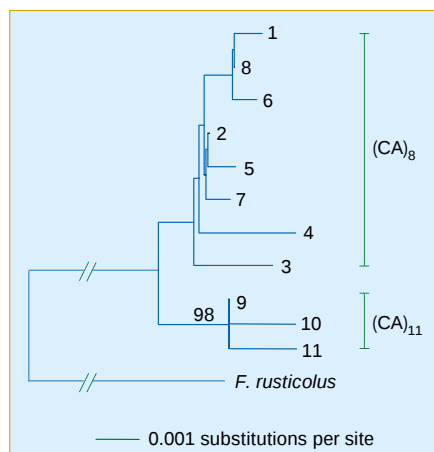
then the genealogical patterns of the W chromosome and mtDNA sequences should be incongruent.

Although polymorphism on avian W chromosomes is generally very low<sup>6</sup>, a polymorphic, W-linked locus containing tandem-repeat microsatellite DNA, *NVHfp49*, has been identified in the peregrine falcon (*Falco peregrinus*)<sup>7</sup>. In a sample of 53 female peregrine falcons from Europe and North America (provided by P. Lindberg, S. Hamrén and C. Keyser), we found two different alleles at this locus, with frequencies of 0.58 and 0.42. Sequencing revealed that the two alleles were composed of 8 and 11 (CA)<sub>n</sub> dinucleotide-repeat units, respectively.

The closely related gyr falcon (*Falco rusticolus*) is monomorphic for this locus<sup>7</sup> and was found to be fixed for the (CA)<sub>8</sub> allele, which therefore probably represents the ancestral state of *F. peregrinus*. If so, the (CA)<sub>8</sub> allele has remained unaltered in one W-chromosome lineage in *F. peregrinus*, whereas it has undergone three steps of repeat expansion in another.

Screening of DNA-sequence variation in 1,625 base pairs of mtDNA of the same 53 female birds uncovered 18 segregating sites, of which 7 were parsimoniously informative (Table 1). Altogether, we identified 11 different haplotypes. A phylogenetic tree constructed from the mtDNA sequence data, rooted with a sequence from *F. rusticolus*, revealed two distinct and basal haplotypic lineages, one of which was resolved with strong bootstrap support. This clade consisted of haplotypes 9–11 (Fig. 1).

All 22 individuals with haplotypes 9–11



**Figure 1** Relationship between mitochondrial (mt) DNA and W-chromosome lineages. A neighbour-joining phylogram on the basis of *Falco peregrinus* mtDNA sequences is shown, rooted with a sequence from *Falco rusticolus* and constructed using PAUP 4.0b3a (Sinauer Associates); numbers are bootstrap values. The association with microsatellite alleles on the W chromosome is shown on the right. The mtDNA sequence analysed corresponds to nucleotide positions 7,158–7,857 (cytochrome oxidase II gene) and 14,512–16,480, with the exception of nucleotides 14,968–15,171 and 15,661–15,704 (control region); GenBank accession number, AF090338.

**Table 1** Summary of polymorphism data for *Falco peregrinus* mitochondrial DNA

| Haplotype | Nucleotide position |       |       |       |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|-----------|---------------------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|           | 7,278               | 7,569 | 7,574 | 7,793 | 14,834 | 14,858 | 14,961 | 15,217 | 15,189 | 15,331 | 15,332 | 15,390 | 15,536 | 15,734 | 15,792 | 15,882 | 16,070 | 16,072 |
| 1         | C                   | A     | G     | T     | C      | G      | C      | A      | C      | T      | C      | A      | A      | A      | A      | G      | C      | A      |
| 2         | -                   | -     | -     | -     | -      | A      | -      | -      | -      | -      | T      | -      | -      | -      | -      | -      | -      | -      |
| 3         | T                   | -     | -     | -     | -      | A      | -      | -      | T      | -      | T      | G      | -      | -      | -      | -      | -      | -      |
| 4         | -                   | -     | -     | C     | -      | A      | -      | -      | T      | -      | T      | -      | G      | -      | G      | -      | -      | -      |
| 5         | -                   | -     | -     | -     | C      | A      | -      | -      | -      | -      | T      | -      | -      | -      | -      | -      | -      | -      |
| 6         | -                   | -     | -     | -     | -      | A      | -      | -      | -      | -      | -      | -      | -      | -      | G      | -      | -      | -      |
| 7         | -                   | -     | -     | -     | -      | A      | -      | -      | -      | -      | T      | -      | -      | -      | -      | A      | -      | -      |
| 8         | -                   | -     | -     | -     | -      | A      | -      | -      | -      | -      | -      | -      | -      | -      | -      | -      | -      | -      |
| 9         | -                   | -     | A     | -     | -      | A      | T      | -      | -      | C      | T      | -      | -      | -      | -      | -      | -      | G      |
| 10        | -                   | -     | A     | -     | -      | A      | T      | -      | -      | C      | T      | -      | -      | -      | -      | -      | T      | G      |
| 11        | -                   | G     | A     | -     | -      | A      | T      | T      | -      | C      | T      | -      | -      | -      | -      | -      | -      | G      |

Nucleotides are shown for haplotype 1 ('master' sequence) at the indicated positions in the whole mitochondrial genome sequence of *F. peregrinus*. In the other haplotypes, only those nucleotides that differ from the master sequence at these positions are shown.

carried the (CA)<sub>11</sub> allele on the W chromosome. Moreover, all 31 individuals with haplotypes 1–8, which formed the other clade, carried the (CA)<sub>8</sub> allele. Thus, the patterns of divergence of mtDNA and W-chromosome sequences are completely concordant. There is nothing in these data to suggest that recombination has occurred between mtDNA haplotypes from the two principal clades since the divergence of the two W-chromosome lineages defined by these microsatellites.

The time over which clonal co-transmission of mtDNA and W-chromosome lineages has taken place in *F. peregrinus* can be estimated by applying a 'molecular clock' to our data. For the mtDNA control region, which is known to evolve faster than other mtDNA sequences, the sequences of the haplotypes from the two principal mtDNA clades diverged by  $0.54 \pm 0.19\%$ . Assuming a divergence rate of 2% per million years, a figure that is commonly applied in studies of avian control-region sequences<sup>8</sup>, this suggests that the two mtDNA clades separated roughly 270,000 years, or 55,000 generations<sup>9</sup>, ago. Alternatively, by using the internal split of haplotypes 9–11 as a conservative minimum time without recombination, they may have separated 110,000 years, or 22,000 generations, ago.

Estimating the divergence of microsatellite alleles is less straightforward because mutation rates vary between loci<sup>10</sup>. However, given the inverse relationship between microsatellite repeat number and mutation rate<sup>10</sup>, and the fact that *NVHfp49* alleles are comparatively short, a mutation rate of less than 1 per  $10^4$  meioses is realistic. Using ( $d\mu$ )<sup>2</sup> statistics<sup>11</sup>, a model for microsatellite evolution, gives a divergence time of over 225,000 years, more than 45,000 generations. We conclude that the transmission of

mtDNA and of W-chromosome lineages in *F. peregrinus* have been concordant over an appreciable evolutionary timescale.

The view that mtDNA in higher animals is clonally inherited only from mothers to daughters, without recombination, has been challenged by data on homoplasmy and linkage disequilibrium in human and other primate mtDNA sequences<sup>1,2</sup>. However, these conclusions have been criticized on methodological grounds<sup>12</sup> and are not supported by new data<sup>13,14</sup>. This is, to our knowledge, the first comparison of the phylogenetic signal of mtDNA sequences with that of a nuclear, female-specific, truly clonal marker, and we conclude that mtDNA recombination, if it occurs, has not been sufficiently common to cast doubt on the standard paradigm of clonal inheritance in this avian model.

**Sofia Berlin, Hans Ellegren**

Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36, Uppsala, Sweden  
e-mail: hans.ellegren@ebc.uu.se

- Awadalla, P., Eyre-Walker, A. & Maynard Smith, J. *Science* **286**, 2524–2525 (1999).
- Eyre-Walker, A., Smith, N. H. & Maynard Smith, J. *Proc. R. Soc. Lond. B* **266**, 477–483 (1999).
- Ellegren, H. *Trends Ecol. Evol.* **15**, 188–192 (2000).
- Fridolfsson, A.-K. & Ellegren, H. *Genetics* **155**, 1903–1912 (2000).
- García-Moreno, J. & Mindell, D. P. *Mol. Biol. Evol.* **17**, 1826–1832 (2000).
- Montell, H., Fridolfsson, A. K. & Ellegren, H. *Mol. Biol. Evol.* (in the press).
- Nesje, M. & Roed, K. H. *Hereditas* **132**, 261–263 (2000).
- Klicka, J. & Zink, R. M. *Science* **277**, 1666–1669 (1997).
- Cramp, S. & Simmons, K. E. L. (eds) *Handbook of the Birds of Europe, the Middle East and North Africa* Vol. 2 (Oxford Univ. Press, Oxford, 1980).
- Ellegren, H. *Trends Genet.* **16**, 551–558 (2000).
- Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L. & Feldman, M. W. *Proc. Natl Acad. Sci. USA* **92**, 6723–6727 (1995).
- Kumar, S., Hedrick, P., Dowling, T. & Stoneking, M. *Science* **288**, 1931 (2000).
- Ingman, M., Kaessman, H., Pääbo, S. & Gyllenstein, U. *Nature* **408**, 708–713 (2000).
- Elson, J. et al. *Am. J. Hum. Genet.* **68**, 145–153 (2001).

**brief communications** is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website ([www.nature.com](http://www.nature.com)) or on request from [nature@nature.com](mailto:nature@nature.com)

# Prokaryotic origin of the actin cytoskeleton

Fusinita van den Ent, Linda A. Amos & Jan Löwe

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

**It was thought until recently that bacteria lack the actin or tubulin filament networks that organize eukaryotic cytoplasm. However, we show here that the bacterial MreB protein assembles into filaments with a subunit repeat similar to that of F-actin—the physiological polymer of eukaryotic actin. By elucidating the MreB crystal structure we demonstrate that MreB and actin are very similar in three dimensions. Moreover, the crystals contain protofilaments, allowing visualization of actin-like strands at atomic resolution. The structure of the MreB protofilament is in remarkably good agreement with the model for F-actin, showing that the proteins assemble in identical orientations. The actin-like properties of MreB explain the finding that MreB forms large fibrous spirals under the cell membrane of rod-shaped cells, where they are involved in cell-shape determination. Thus, prokaryotes are now known to possess homologues both of tubulin, namely FtsZ, and of actin.**

A central component of the eukaryotic cytoskeleton is filamentous actin. Actin is the most abundant protein in many eukaryotic cells, and is conserved from yeast to humans<sup>1</sup>. In 1942, Straub isolated monomeric actin (G-actin) and discovered that raising the salt concentration causes G-actin to polymerize into filamentous actin (F-actin)<sup>2</sup>. Electron microscopy and X-ray fibre diffraction have shown that F-actin consists of two protofilaments that are twisted gently around one another to form a right-handed double helix. The subunits in each actin protofilament have an approximately 55 Å spacing<sup>3</sup>, and the helical pitch is variable owing to torsional flexibility of the protofilaments<sup>4</sup>. Under appropriate conditions actin will polymerize into a variety of polymers<sup>5</sup>. When actin is treated with gadolinium it assembles into sheets and cylinders of straight protofilaments<sup>6</sup>. The filamentous polymers of actin determine the shape of many eukaryotic cells, besides having a vast range of other functions. The three-dimensional structure of G-actin—which has a relative molecular mass of 43,000 ( $M_r$  43K)—has been determined in complexes with various actin-binding proteins that prevent actin polymerization<sup>7–10</sup>.

Actin is a member of a larger superfamily of proteins<sup>11,12</sup>, which includes Hsp70 (ref. 13), cell-division protein FtsA<sup>14</sup>, and sugar kinases<sup>15,16</sup>. Crystal structures have revealed that each member of the actin superfamily has the characteristic core of actin, and is distinguished by additional insertions or deletions that are necessary for the specific function of each family member. A sequence database search revealed that the bacterial proteins MreB and StbA have sequence patterns in common with the actin superfamily<sup>11</sup>. MreB, among all proteins of the superfamily, is most closely related to actin in overall size<sup>11</sup>.

The *mreB* gene is located in the gene cluster *mre* (murein cluster e). It is, together with *mrd*, the principal operon involved in determination of cell shape in bacteria<sup>17–19</sup>. Recent evidence shows that MreB assembles into a cytoskeleton-like structure in *Bacillus subtilis*<sup>20</sup>. MreB and the closely related protein Mbl are important in regulating the cell shape of *B. subtilis*; immunofluorescence reveals that elongated polymers of MreB and Mbl encircle the cell as large spirals under the cell membrane. The MreB proteins are widely distributed among rod-shaped, filamentous and helical bacteria<sup>20</sup>, suggesting that an MreB cytoskeleton is important to generate a non-spherical shape.

To investigate whether MreB can self assemble into actin-like filaments, we cloned and purified MreB from *Thermotoga maritima*. Biochemical and electron microscopic analyses show that the protein forms filaments with a longitudinal repeat similar to that

of eukaryotic actin. Elucidation of the crystal structure of MreB shows that it is indeed clearly related to actin. Furthermore, the crystal packing reveals the filamentous structure of an actin-like protein at atomic resolution. Here we provide biochemical and structural evidence for MreB as the bacterial actin homologue.

## Polymerization assays

The gene encoding MreB1 from *T. maritima* was amplified by polymerase chain reaction (PCR) and cloned for overexpression in *Escherichia coli* strain C41 (see Methods). *Thermotoga maritima* has two *mreB* genes. We also cloned *mreB2*, but the protein was mostly insoluble (data not shown). A BLAST search with both proteins revealed that MreB1 is more closely related to MreB from *B. subtilis*, with 56% overall identity.

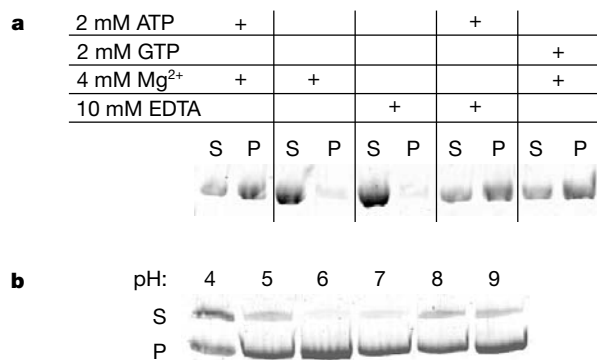
MreB forms polymers under various conditions, as was initially investigated in a pelleting assay (Fig. 1). To polymerize, MreB requires ATP (Fig. 1a, lane 1) or GTP (lane 5). It can form filaments in the absence of magnesium (lane 4). Polymers are formed over a wide pH range, the optimum being pH 6–7 (Fig. 1b). In contrast to actin polymerization, which requires physiological salt concentrations, *T. maritima* MreB is able to form filaments over a wide range of salt concentrations, as high as 4 M NaCl. The fact that thermophilic organisms usually possess relatively high intracellular salt concentrations could explain the ability of *T. maritima* MreB to assemble in high salt.

## Electron microscopy

The nature of MreB polymers, found in the pelleting assay, was investigated by electron microscopy of negatively stained samples. A variety of polymers formed under different conditions (see Fig. 2a, e). The simplest polymers are thin filaments that appear to consist of two protofilaments (each composed of a string of monomers), but such individual thin filaments are rare. More common are pairs of thin filaments, which are often curved (Fig. 2a) depending on the conditions used. Those shown in Fig. 2a are much more highly curved than would be required to produce the curved filaments observed *in vivo*<sup>20</sup>. From our images it is unclear how the curvature is accommodated into the structure.

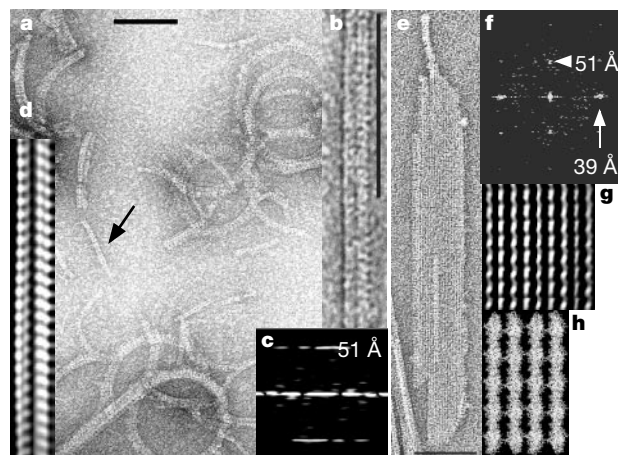
The filtered image (Fig. 2d) of the polymer in Fig. 2b (diameter of about 160 Å) apparently shows two thin filaments, with an approximately 51 Å longitudinal spacing (Fig. 2c). The thin filaments appear to lack the distinct twist of the two-stranded helices of F-actin, although they often appear to have a slight twist. At low NaCl concentrations (25 mM NaCl), MreB forms two-dimensional





**Figure 1** MreB polymerization assays. MreB was incubated at 37 °C with nucleotide and NaCl, and was centrifuged at 140,000g. S, supernatant; P, solubilized pellet. Proteins were separated on a 12% SDS gel and stained with PAGE 83. **a**, Effect of nucleotides and magnesium. MreB requires ATP or GTP, but not magnesium, for polymerization. **b**, pH optimum. MreB was incubated under standard conditions (see Methods) with 100 mM buffer: citrate pH 4, pH 5, MES pH 6, HEPES pH 7, Tris-HCl pH 8, BICINE pH 9. Maximal pelleting was observed at pH 6–7.

crystalline sheets (Fig. 2e). The diffraction pattern from the sheets also shows a strong layer line at 51 Å, and in addition a spacing of 39 Å in the equatorial plane (Fig. 2f). Filtered images of the sheets (Fig. 2g) show individual filaments that are consistent with the protofilaments in the crystal structure (see below). They seem to be related to the gadolinium-induced sheets of actin, which consist of untwisted protofilaments arranged in an antiparallel



**Figure 2** Electron micrographs of negatively stained MreB filaments. **a**, Typical view of MreB filaments when salt and neutral pH is used. MreB (1 mg ml<sup>-1</sup>) was incubated for 30 min at 37 °C in 100 mM Tris-HCl, pH 7.5, 100 mM NaCl, 3 mM CaCl<sub>2</sub>, 2 mM ATP (buffered) and 4 mM MgCl<sub>2</sub>. Scale bar, 100 nm. The filament indicated with an arrow is the same filament that is shown at a higher magnification in **b**. **b**, 'Double' filament formed at high pH. MreB (1 mg ml<sup>-1</sup>) was incubated for 30 min at 37 °C in 100 mM BICINE, pH 9.0, 100 mM NaCl, 2 mM ATP (buffered) and 4 mM MgCl<sub>2</sub>. Scale bar, 100 nm. **c**, Diffraction image of the polymer in **b** showing the first strong layer line at 51 Å. **d**, Filtered image of **b**, treating both filaments separately. The polymer is about 160 Å wide, which suggests four single protofilaments in total (each 40 Å, see Fig. 5). **e**, Electron micrograph of a negatively stained MreB sheet. MreB (1 mg ml<sup>-1</sup>) was incubated for 30 min at 37 °C in 100 mM Tris-HCl, pH 7.5, 25 mM NaCl, 2 mM ATP (buffered) and 4 mM MgCl<sub>2</sub>. The protofilaments are aligned vertically in the MreB sheet. Scale bar, 100 nm. **f**, Diffraction image of sheet in **e**. The longitudinal repeat is 51 Å, the lateral spacing is 39 Å. **g**, Filtered image of sheet in **e**. **h**, The protofilaments found in the crystals of MreB fit well with the filtered image in **g**. The longitudinal repeat of 51 Å in the sheets is the same as in the crystals (Fig. 5). The lateral spacing in the sheets suggests that the protofilaments interact with their flat sides.

conformation<sup>6</sup>. It is possible, but not yet certain, that the MreB sheets also have an antiparallel arrangement of protofilaments.

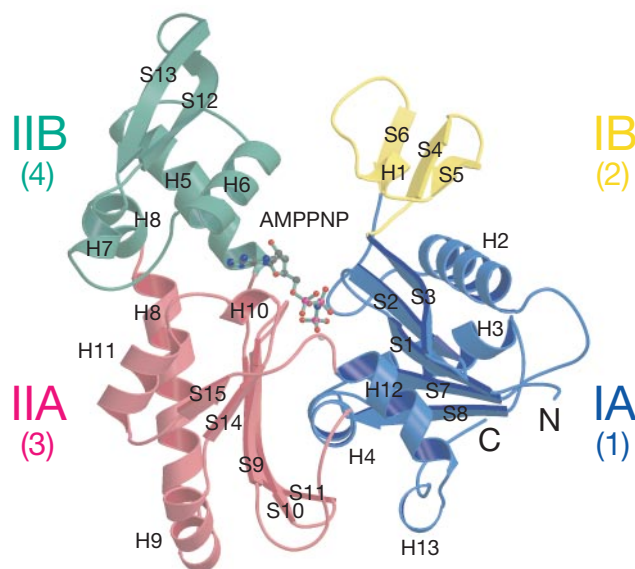
The electron microscopic analysis shows that the polymers found in the pelleting assay and in all other tested conditions have a longitudinal spacing of 51 Å, which is reminiscent of the 55 Å spacing of F-actin's helical strands<sup>3</sup>.

### Crystal structure of MreB

To investigate the homology of MreB to actin in atomic detail, we determined the three-dimensional structure of MreB. We obtained trigonal and monoclinic crystals of MreB from *T. maritima*. The trigonal crystals, space group *P*3<sub>1</sub>21, diffracted to 2.1 Å but were partially twinned. The structure was solved by multiple anomalous dispersion (MAD), using the monoclinic crystals (space group *C*2, see Methods and Supplementary Information Table 1). The structures of both the apo- (NATI) and AMPPNP-containing trigonal crystals were solved by molecular replacement (Table 1).

The structure of MreB confirms that it is a member of the actin family of proteins, and is characterized by two domains (I and II) that hold a nucleotide-binding site in the interdomain cleft<sup>12</sup>. Each domain is subdivided into two subdomains (A and B) (Fig. 3). The two larger subdomains (IA and IIA) have a common fold that comprises a five-stranded β-sheet surrounded by three α-helices. These subdomains are connected through a helix, H4. The smaller subdomains (IB and IIB) are more diverse in the actin superfamily, and are probably unique to the specific function of the proteins. Notably, MreB shows exactly the same topology as actin in the structure of these domains (Fig. 4). This is in contrast to Hsp70, FtsA and, in particular, the sugar kinases. The position where the smaller subdomains are inserted is identical in the two proteins, that is, after the third strand (S3 and S11, respectively) of the β-meander of both subdomain IA and IIA. One significant difference between actin and MreB is a loop that is inserted in actin in the equivalent of helix H8. The loop has been implicated in the intermolecular interaction between the subunits during initial assembly of the filament<sup>3</sup>.

A three-dimensional structural similarity search revealed that MreB can be superimposed on actin (Protein Data Bank (PDB)



**Figure 3** Ribbon representation of the crystal structure of MreB complexed with AMPPNP and magnesium. MreB is a member of the actin family of proteins, showing the typical four-domain architecture. AMPPNP binds in a cleft between domains I and II. The four subdomains IA, IB, IIA and IIB correspond to subdomains 1, 2, 3 and 4 in actin. Virtually no differences are detectable from the two non-ligand-bound crystal structures (HRES, NATI). Images were created with MOLSCRIPT and RASTER3D<sup>38,39</sup>.

**Table 1 Refinement statistics**

|                        | HRES                | NATI                | AMPPNP              |
|------------------------|---------------------|---------------------|---------------------|
| Residues               | 4–336               | 1–336               | 1–335               |
| Water, cofactor        | 266                 | 318                 | 0, AMPPNP           |
| Resolution             | 2.1 Å               | 2.1 Å               | 3.1 Å               |
| Twinning fraction*     |                     | <0.01               | 0.055               |
| R-factor, R-free†      | 0.197, 0.234        | 0.194, 0.240        | 0.206, 0.276        |
| B average‡             | 28.6 Å <sup>2</sup> | 30.3 Å <sup>2</sup> | 25.2 Å <sup>2</sup> |
| Geometry bonds/angles§ | 0.006 Å, 1.233°     | 0.005 Å, 1.264°     | 0.009 Å, 1.421°     |
| Ramachandran           | 91.5%/0%            | 94.1%/0%            | 86.0%/0%            |
| PDB ID¶                | 1JCE                | 1JCF                | 1JCG                |

\* Twinning fraction as used in refinement, operator –h, –k, l.

† Five per cent of reflections were randomly selected for determination of the free R-factor (taking twinning into account where appropriate) before any refinement.

‡ Temperature factors averaged for all atoms.

§ r.m.s. deviations from ideal geometry for bond lengths and restraint angles<sup>36</sup>.

|| Percentage of residues in the ‘most-favoured region’ of the Ramachandran plot and percentage of outliers (PROCHECK<sup>37</sup>).

¶ Protein Data Bank identifiers for coordinates.

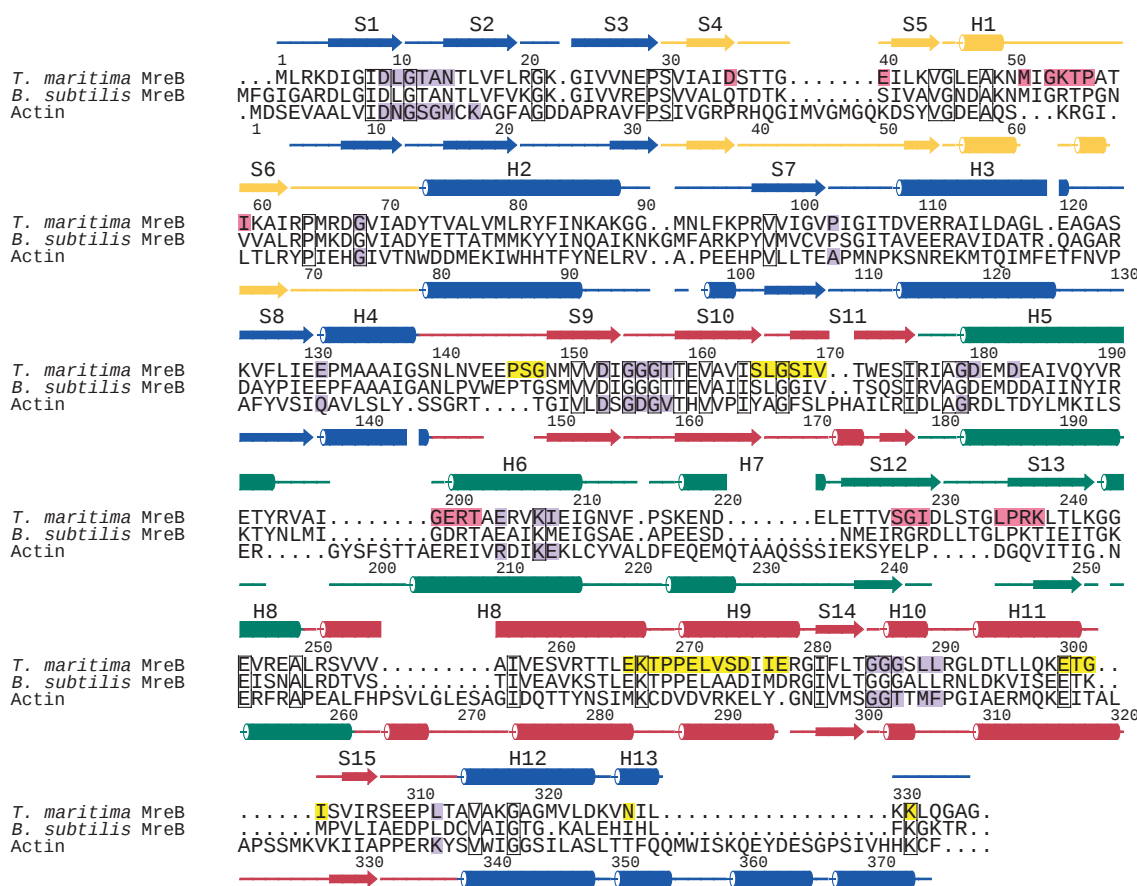
entry 1ATN) with an r.m.s. deviation of 3.7 Å over 310 Cα atoms and a z-score of 29.2 (ref. 21). MreB and HSC70 (constitutive Hsp70 protein; PDB entry 1HPM) superimpose with an r.m.s. deviation of 3.4 Å over 299 Cα atoms and a z-score of 28.8. This gives the impression that MreB is closely related to Hsp70; however, a detailed comparison between the two structures reveals that Hsp70 has a substantial insertion of about 40 residues in subdomain IB, which is absent in MreB and actin. Moreover, Hsp70 contains a principal substrate-binding domain of more than 250 residues at

the carboxy-terminus. The other members of the actin superfamily are more distinct: hexokinase (PDB entry 1BDG) has an r.m.s. deviation of 3.5 Å over 253 equivalent residues and a z-score of 14.3 when compared to MreB. Bacterial cell-division protein FtsA (PDB entry 1E4G) can be superimposed on MreB with an r.m.s. deviation of 2.8 Å over 270 residues and a z-score of 25.2. The fact that FtsA, as a member of the actin family, acts together with the tubulin homologue FtsZ in bacterial cytokinesis suggests that it may also form filaments. However, FtsA does not polymerize into actin-like filaments *in vitro*, as tested under numerous conditions (data not shown). FtsA is distinguished from the other members of the actin family by the position of subdomain IB, which is located on the opposite side of subdomain IA (ref. 14). Another bacterial protein, StbA, which is involved in plasmid segregation is a putative member of the actin superfamily, but is shorter and probably lacks subdomain IIB (ref. 11).

Elucidation of the structure reveals that MreB is most closely related to actin among all members of the actin family. The topology of the secondary structural elements is shared between the two proteins, even in the variable domains IB and IIB (Fig. 4).

### Nucleotide-binding site

The nucleotide and the high-affinity, divalent cation-binding sites of MreB are located near the base of the left between domains I and II. The superposition of the nucleotide in the MreB structure (AMPPNP) on that of the actin structure (ATP; PDB entry 1YAG) shows that most of the active-site residues are at similar positions,



**Figure 4** Sequence alignment of MreB1 from *T. maritima* to *B. subtilis* MreB (SWISSPROT: MREB\_BACSU) and structure-based sequence alignment to yeast actin (SWISSPROT: ACT\_YEAST; PDB, 1YAG) using DALI<sup>21</sup>. Secondary structural elements are shown according to DSSP output of MreB and actin. Helix and strand colours refer to the domain colours in Figs 3 and 5. Active site residues in MreB and actin are highlighted in

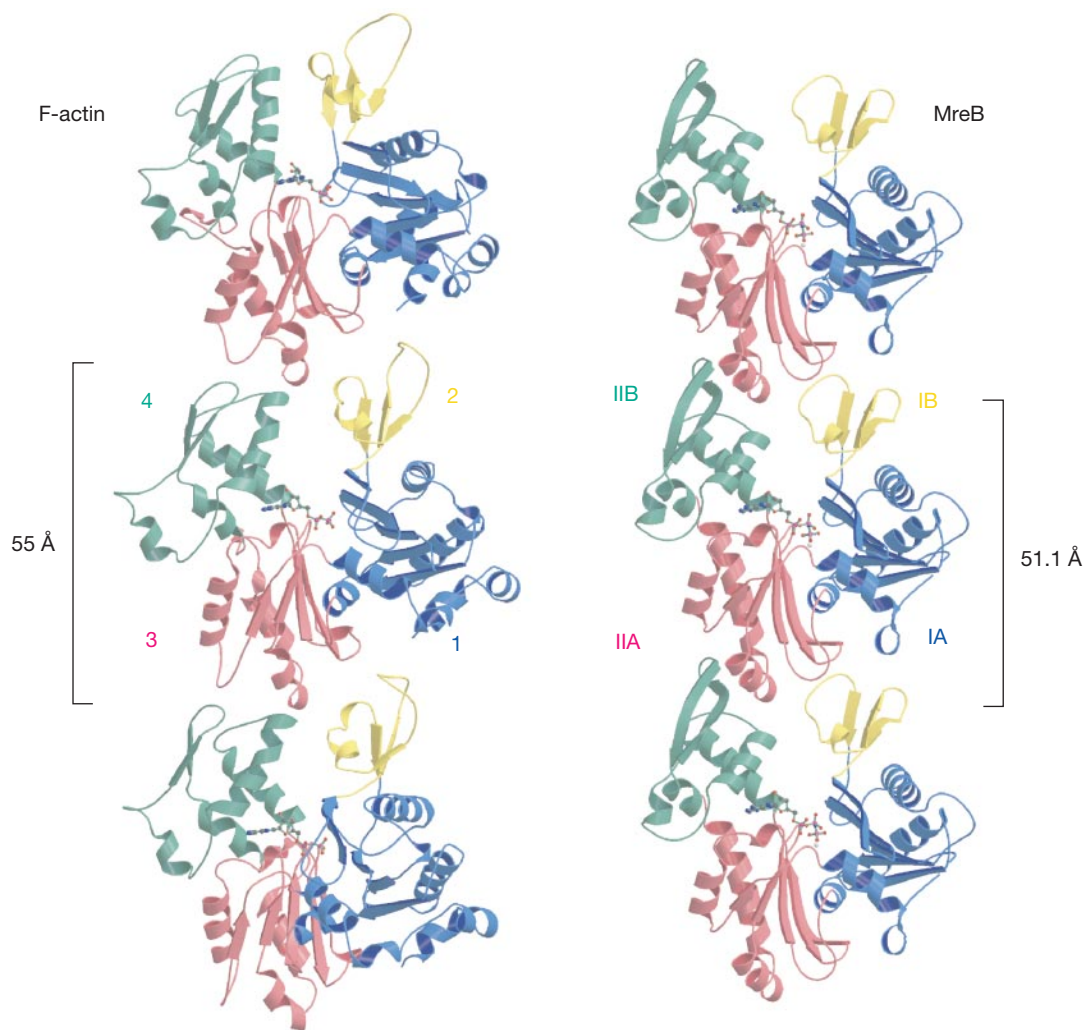
light purple. Protofilament contacts from one subunit to the next are marked in yellow; residues involved in protofilament contacts to the previous subunit are coloured pink (see residues in Fig. 6). Boxed residues are conserved in all three sequences. Sequence identity is 56% between the two MreB proteins and 15% between MreB and actin. The alignment was prepared with ALSCRIPT<sup>40</sup>.

with the exception of the residues located in subdomain IB. These residues, which coordinate the  $\gamma$ -phosphate, are tilted compared with those in actin (see stereo figure in Supplementary Information). This may indicate a different nucleotide state of the two proteins. Most of the residues involved in the coordination of the nucleotide are similar and can be superimposed (purple shaded residues in Fig. 4). Sometimes, where amino acids have changed as in the first phosphate-binding loop (residues 9–14), the interactions are through the amide group of the main chain. Elsewhere, when the residues are different, their role is conserved, as in the hydrophobic pocket that surrounds the adenosine. In actin, the pocket is closed at the top by the hydrophobic part of Glu 214. MreB has Ile 208 at the equivalent position, which has the same role. In addition, actin has a salt bridge between Glu 214 and Arg 210. At the equivalent position of Arg 210, MreB has Glu 204, which forms a salt bridge to Lys 49. This is another example of coevolution, previously reported in a comparison between actin and Hsp70 (ref. 22). The phosphate moiety is bound as in actin, with glycines preceding helix H10 binding to the phosphates. The  $\beta$ -phosphate forms hydrogen bonds with the amides of Ala 13 and Asn 14. The hydrogen bond in actin between Lys 18 and the  $\beta$ -phosphate is not present in MreB, which has a Leu at the equivalent position.

### Crystal structure of the MreB protofilament

A notable feature of the crystal packing of both the monoclinic and the trigonal crystal forms is that they contain an MreB protofilament in which the subunits are translated in one dimension (see Supplementary Information for the coordinates of the protofilament). The straight nature of the protofilament is an ideal building block for crystals, which is reflected by the fact that crystals were obtained under many conditions. Most importantly, the spacing between the monomers in the longitudinal direction is identical to the spacing deduced from the electron micrographs (51 Å, Fig. 5). The presence of these protofilaments in both crystal forms and in all of the polymers that were investigated under the electron microscope indicates a high propensity of MreB to form protofilaments. It also proves that intra-protofilament contacts are biologically relevant and not just due to crystal packing.

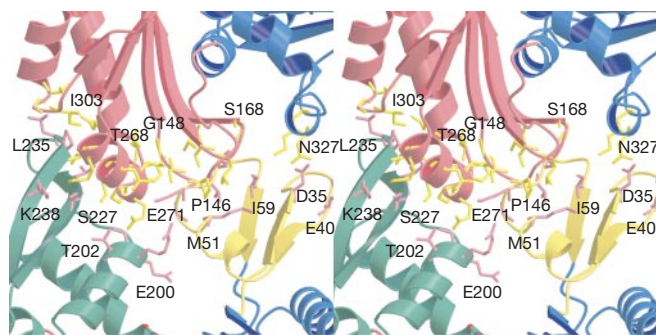
A closer look at the interface that is formed by two monomers along a protofilament reveals that it is primarily composed of hydrophobic residues (Fig. 6). The two strands S12 (residues 227–229) and S13 (residues 235–238) interact with helix H9 (residues 266–275, 277, 278) and with the loop between helix H11 and strand S15 (residues 300–303). Helix H9 also makes contacts with helix H6 (residues 199–202). The protofilament interface is extended by hydrophobic interactions between residues



**Figure 5** Both crystal forms (HRES, NATI) contain one-dimensional protofilaments. Residues at the bottom of subdomains IA and IIA (1 and 3 in actin) insert into the cleft formed by subdomains IB and IIB (2 and 4 in actin). This is basically the same interaction that has been proposed for the longitudinal interaction in the two strands of F-actin, shown

in the left panel<sup>3</sup>. The protofilament repeat is 51.1 Å as measured in crystal form NATI. In the trigonal crystal form the protofilament axis is aligned with the crystallographic cell axis *a*.





**Figure 6** Close-up of the protofilament protein interface. The interaction surface measures  $1,709 \text{ \AA}^2$ , equalling 11.8% of the solvent-accessible surface of monomeric MreB. Residues are coloured as in Fig. 4 with yellow for those involved in contacts from one subunit to the one above, and pink for contacts to the subunit below. The image was created with MOLSCRIPT and RASTER3D<sup>38,39</sup>.

from the loop connecting helix H1 with strand S6 (residues 51–59) and the two loops preceding strands S9 (residues 146–148) and S11 (residues 165–170). Hydrophilic interactions between Lys 330 and Asp 40, as well as Lys 331 and Asp 35, complete the interface.

The resemblance between actin and MreB is reflected in the structure of the filament as well as in the longitudinal repeat (Fig. 5). In addition, actin and MreB are in almost identical orientation in the protofilament, resulting in the same longitudinal contact region. A discrepancy between F-actin and the filaments of MreB is that F-actin is a twisted pair of protofilaments. It is possible that the twist is imposed by different lateral interactions in actin, such as those from the actin-specific loop inserted in H8. In any case, the lateral interactions in actin are variable as the twisted parallel arrangement of the protofilaments in F-actin can be converted into the straight antiparallel organization in the actin sheets<sup>6</sup>. The filtered image of an MreB sheet (Fig. 2g) can also be explained by a side view of adjacent protofilaments (Fig. 2h). At the present resolution, the protofilaments do not show a definite polarity; adjacent protofilaments in the sheets are not obviously different. Cryo-electron microscopy is needed for a detailed analysis of the arrangement of the protofilaments in both sheets and filaments.

We present here an actin-like protofilament in atomic detail. Because the current model of F-actin is of low resolution, a detailed comparison between the protofilament interface of MreB and actin is not possible at the moment. Interestingly, the refined model of F-actin<sup>23,24</sup> is less similar to our crystal structure than the initial unrefined model, in which the crystal structure of G-actin in complex with DNase I was used<sup>3</sup>. This finding is supported by the fact that the significant conformational changes in the refined F-actin model are inconsistent with the structural changes identified in the analysis of domain movements made on the basis of four different crystal structures of actin<sup>25</sup>.

Although it is impossible to compare the protofilament interface in detail, the structure-based sequence alignment between MreB and actin shows that the residues involved in the monomer contacts along the protofilament are not well conserved (Fig. 4). Apparently, evolution has allowed for (concomitant) mutations of residues on both sides of the interface. A low degree of conservation in the protofilament interface has also been found for filaments of FtsZ and tubulin<sup>26</sup>.

Previous studies implicated MreB in the determination of cell shape in bacteria and showed MreB-dependent cytoskeletal structures<sup>20</sup>, suggesting a functional similarity between MreB and actin. To further support this relationship, future experiments are needed, which may shed light on the dynamic behaviour of MreB. It remains to be seen whether MreB shares other properties with actin

such as treadmilling, or forming a track for motor proteins. Nevertheless, the biochemical and structural data presented here strongly suggest that MreB is an excellent candidate for the actin homologue in bacteria. □

## Methods

### Protein expression and purification

The two MreB homologues from *T. maritima* (DSMZ number 3109: MreB1, TM0588 (SWALL: Q9WZ57); MreB2, TM1544 (SWALL: Q9X1N0)) were amplified by genomic PCR using primers that introduce unique cleavage sites for *NdeI* and *BamHI* into the product. The fragments were cloned into pHis17. This enables the proteins to be expressed under the control of the T7 promoter and adds eight residues to the C-terminus (GSHHHHHH). C41(DE3) cells<sup>27</sup> were transformed, grown at 37 °C and induced in log phase ( $A_{600} = 0.6$ ) with 1 mM IPTG. The cells were collected after 4 h induction, and the pellets were frozen in liquid nitrogen. Cells were thawed and lysed by the addition of lysozyme and subsequent sonication, in a buffer of 50 mM Tris-HCl pH 8.0. DNase I was added and after centrifugation (140,000g) the lysate was applied to a  $\text{Ni}^{2+}$ -NTA column (QIAGEN). Buffer A was 300 mM NaCl and 50 mM Tris-HCl pH 6.0; buffer B was 1 M imidazole and 50 mM Tris-HCl, pH 6.0. After washing with 5% buffer B the protein was eluted with 30% buffer B. Peak fractions were pooled and concentrated before loading onto a sephacryl S200 column (Amersham-Pharmacia) equilibrated in 20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 1 mM  $\text{NaN}_3$ , pH 7.5.

### Polymerization assays

Purified *T. maritima* MreB (35.8K) was mixed in a volume of 40  $\mu\text{l}$  with 100 mM Tris-HCl, pH 7.0, 2 mM nucleotides, 4 mM  $\text{MgCl}_2$  and 100 mM NaCl, and incubated at 37 °C for 30 min, if not stated otherwise. Final protein concentration was  $1 \text{ mg ml}^{-1}$  (28  $\mu\text{M}$ ). (For exact conditions please refer to legend of Fig. 1a, b.) The reactions were centrifuged at 140,000g in a Beckman TLA100 rotor for 20 min at 20 °C. The supernatant was removed for analysis and the pellet washed with buffer containing the same components as the reaction buffer minus protein. Pellets were solubilized with SDS-containing gel loading buffer, using the same volume as for the supernatant. Samples were analysed on 12% SDS–polyacrylamide gel electrophoresis (PAGE) gels.

### Electron microscopy

MreB at  $1 \text{ mg ml}^{-1}$  in polymerization buffer (see Fig. 2 legend) was incubated at 37 °C for 30 min. Next, 5  $\mu\text{l}$  of the reaction was put on carbon-coated copper electron microscopic grids. Excess liquid was blotted off. Grids were washed with a drop of water, stained with 2% uranyl acetate solution and dried. Pictures were taken with a Philips 208 electron microscope on film plates at  $\times 40,000$ – $\times 80,000$  magnification. Images were digitized on a Zeiss SCAI CCD scanner at 21–28- $\mu\text{m}$  resolution, and analysed using MRC image processing software<sup>28</sup>.

### Crystallization and structural determination

Seleno-methionine (SeMet)-substituted MreB protein was expressed as described<sup>29</sup>. For purification of SeMet-substituted protein, buffers A and B contained 5 mM  $\beta$ -mercapto-ethanol, and the final buffer contained 5 mM dithiothreitol. Native crystals were grown using the sitting-drop vapour diffusion technique using 10% PEG 8000, 200 mM NaCl and 0.1 M CAPS, pH 10.5 as the crystallization solution. Drops composed of 1  $\mu\text{l}$  protein at  $8 \text{ mg ml}^{-1}$  and 1  $\mu\text{l}$  crystallization solution were equilibrated for a minimum of 3 days at 19 °C. Seleno-methionine-substituted crystals were grown in the same manner as for the native protein but with 10% isopropanol, 200 mM NaCl and 0.1 M HEPES, pH 7.5. Crystals were frozen in mother liquor complemented with 25% PEG400 or 30% glycerol for the native and SeMet crystals, respectively. Crystals belong to space group P3<sub>1</sub>21 (native) or C2 (SeMet). Two SeMet MAD datasets and the native dataset were collected at ID14-4, ESRF, Grenoble (see Supplementary Information Table 1). Crystals were indexed and integrated using the MOSFLM package<sup>30</sup>, and data were further processed using the CCP4 package<sup>31</sup>.

An initial 2.4 Å MAD density map was generated by locating eight selenium sites using the program SOLVE<sup>32</sup>, which was also used to calculate phases. Final figure of merit of the phases was 0.7 for all data to 2.4 Å resolution. We used RESOLVE<sup>33</sup> for solvent flattening, assuming 60% solvent content. All ordered residues were built into the MAD electron density map using MAIN<sup>34</sup>. The structure was refined against all data in dataset HRES (SeMet inflection point) to 2.1 Å resolution using CNS<sup>35</sup>, taking anomalous and dispersive terms for the selenium atoms into account. Dataset NATI showed weak twinning when comparing cumulative intensity distributions with those from randomly scattered atoms (TRUNCATE). However, the twinning fraction was found to be less than 1% after refinement, and was ignored. NATI was solved by molecular replacement using the refined HRES model and AMORE. Dataset AMPPNP was collected in-house from a trigonal crystal soaked for 2–3 h in mother liquor supplemented with 2 mM AMPPNP and 4 mM  $\text{MgCl}_2$ . The crystal was isomorphous to NATI and the structure was solved by difference Fourier methods, clearly showing the difference density for AMPPNP and magnesium. Details of the refined models are summarized in Table 1.

Received 13 June; accepted 2 August 2001.

- Schmidt, A. & Hall, M. N. Signaling to the actin cytoskeleton. *Annu. Rev. Cell Dev. Biol.* **14**, 305–338 (1998).
- Straub, F. B. Actin. *Stud. Inst. Med. Chem. Univ. Szeged* **2**, 3–15 (1942).

3. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).
4. Egelman, E. H., Francis, N. & DeRosier, D. J. F-actin is a helix with a random variable twist. *Nature* **298**, 131–135 (1982).
5. Steinmetz, M. O. *et al.* An atomic model of crystalline actin tubes: combining electron microscopy with X-ray crystallography. *J. Mol. Biol.* **278**, 703–711 (1998).
6. Aebi, U., Fowler, W. E., Isenberg, G., Pollard, T. D. & Smith, P. R. Crystalline actin sheets—their structure and polymorphism. *J. Cell Biol.* **91**, 340–351 (1981).
7. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic-structure of the actin–DNase-I complex. *Nature* **347**, 37–44 (1990).
8. McLaughlin, P. J., Gooch, J. T., Mannherz, H. G. & Weeds, A. G. Structure of gelsolin segment-1-actin complex and the mechanism of filament severing. *Nature* **364**, 685–692 (1993).
9. Schutt, C. E., Myslik, J. C., Rozycki, M. D., Goonesekere, N. C. W. & Lindberg, U. The structure of crystalline profilin  $\beta$ -actin. *Nature* **365**, 810–816 (1993).
10. Chik, J. K., Lindberg, U. & Schutt, C. E. The structure of an open state of  $\beta$ -actin at 2.65 angstrom resolution. *J. Mol. Biol.* **263**, 607–623 (1996).
11. Bork, P., Sander, C. & Valencia, A. An ATPase domain common to prokaryotic cell-cycle proteins, sugar kinases, actin, and Hsp70 heat-shock proteins. *Proc. Natl Acad. Sci. USA* **89**, 7290–7294 (1992).
12. Kabsch, W. & Holmes, K. C. Protein motifs 2. The actin fold. *FASEB J.* **9**, 167–174 (1995).
13. Flaherty, K. M., Delucaflaherty, C. & McKay, D. B. Three-dimensional structure of the ATPase fragment of a 70kDa heat-shock cognate protein. *Nature* **346**, 623–628 (1990).
14. van den Ent, F. & Löwe, J. Crystal structure of the cell division protein FtsA from *Thermotoga maritima*. *EMBO J.* **19**, 5300–5307 (2000).
15. Hurley, J. H. *et al.* Structure of the regulatory complex of *Escherichia coli* Ilii(Glc) with glycerol kinase. *Science* **259**, 673–677 (1993).
16. Anderson, C. M., McDonald, R. C. & Steitz, T. A. Sequencing a protein by X-ray crystallography. Interpretation of yeast hexokinase B at 2.5 Å resolution by model building. *J. Mol. Biol.* **123**, 1–13 (1978).
17. Wachi, M. *et al.* Mutant isolation and molecular cloning of *mre* genes, which determine cell shape, sensitivity to mecillinam, and amount of penicillin-binding proteins in *Escherichia coli*. *J. Bacteriol.* **169**, 4935–4940 (1987).
18. Doi, M. *et al.* Determinations of the DNA sequence of the *mreB* gene and of the gene products of the *mre* region that function in formation of the rod shape of *Escherichia coli* cells. *J. Bacteriol.* **170**, 4619–4624 (1988).
19. Levin, P. A., Margolis, P. S., Setlow, P., Losick, R. & Sun, D. X. Identification of *Bacillus subtilis* genes for septum placement and shape determination. *J. Bacteriol.* **174**, 6717–6728 (1992).
20. Jones, L. J. F., Carballido-Lopez, R. & Errington, J. Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. *Cell* **104**, 913–922 (2001).
21. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
22. Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. Similarity of the 3-dimensional structures of actin and the ATPase fragment of a 70-kDa heat-shock cognate protein. *Proc. Natl Acad. Sci. USA* **88**, 5041–5045 (1991).
23. Lorenz, M., Popp, D. & Holmes, K. C. Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J. Mol. Biol.* **234**, 826–836 (1993).
24. Tirion, M. M., Benavraham, D., Lorenz, M. & Holmes, K. C. Normal-modes as refinement parameters for the F-actin model. *Biophys. J.* **68**, 5–12 (1995).
25. Page, R., Lindberg, U. & Schutt, C. E. Domain motions in actin. *J. Mol. Biol.* **280**, 463–474 (1998).
26. Nogales, E., Downing, K. H., Amos, L. A. & Löwe, J. Tubulin and FtsZ form a distinct family of GTPases. *Nature Struct. Biol.* **5**, 451–458 (1998).
27. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
28. Crowther, R. A., Henderson, R. & Smith, J. M. MRC image processing programs. *J. Struct. Biol.* **116**, 9–16 (1996).
29. van den Ent, F., Lockhart, A., Kendrick-Jones, J. & Löwe, J. Crystal structure of the N-terminal domain of MukB: a protein involved in chromosome partitioning. *Struct. Fold. Des.* **7**, 1181–1187 (1999).
30. Leslie, A. G. W. *Recent Changes to the MOSFLM Package for Processing Film and Image Plate Data* (SERC Laboratory, Daresbury, UK, 1991).
31. Collaborative Computing Project No. 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
33. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
34. Turk, D. *Weiterentwicklung eines Programms für Molekülgrafik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen*. Thesis, Technische Univ. München (1992).
35. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
36. Engh, R. A. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392–400 (1991).
37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
38. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
39. Merritt, E. A. & Bacon, D. J. Raster 3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).
40. Barton, G. J. ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37–40 (1993).

**Supplementary information** is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as a paper copy from the London editorial office of *Nature*.

## Acknowledgements

We would like to thank S. Munro for critically reading the manuscript and the staff at ID14-4 of ESRF (Grenoble, France) for assistance with MAD data collection.

Correspondence and requests for materials should be addressed to F.V.D.E. (e-mail: [fent@mrclmb.cam.ac.uk](mailto:fent@mrclmb.cam.ac.uk)). The coordinates for HRES (space group C2), NATI (space group P3<sub>1</sub>21) and AMPPNP are deposited in the Protein Data Bank under accession numbers 1JCE, 1JCF and 1JCF, respectively.

# Rapid X-ray flaring from the direction of the supermassive black hole at the Galactic Centre

F. K. Baganoff\*, M. W. Bautz\*, W. N. Brandt†, G. Chartas†, E. D. Feigelson†, G. P. Garmire†, Y. Maeda†‡, M. Morris§, G. R. Ricker\*, L. K. Townsley† & F. Walter||

\* Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA

† Department of Astronomy and Astrophysics, Pennsylvania State University, University Park, Pennsylvania 16802-6305, USA

‡ Institute of Space and Astronautical Science, 3-1-1 Yoshinodai, Sagami-hara, 229-8501, Japan

§ Department of Physics and Astronomy, University of California at Los Angeles, Los Angeles, California 90095-1562, USA

|| Department of Astronomy, California Institute of Technology, Pasadena, California 91125, USA

The nuclei of most galaxies are now believed to harbour supermassive black holes<sup>1</sup>. The motions of stars in the central few light years of our Milky Way Galaxy indicate the presence of a dark object with a mass of about  $2.6 \times 10^6$  solar masses (refs 2, 3). This object is spatially coincident with the compact radio source Sagittarius A\* (Sgr A\*) at the dynamical centre of the Galaxy, and the radio emission is thought to be powered by the gravitational potential energy released by matter as it accretes onto a supermassive black hole<sup>4,5</sup>. Sgr A\* is, however, much fainter than expected at all wavelengths, especially in X-rays, which has cast some doubt on this model. The first strong evidence for X-ray emission was found only recently<sup>6</sup>. Here we report the discovery of rapid X-ray flaring from the direction of Sgr A\*, which, together with the previously reported steady X-ray emission, provides compelling evidence that the emission is coming from the accretion of gas onto a supermassive black hole at the Galactic Centre.

Our view of Sgr A\* in the optical and ultraviolet wavebands is blocked by the large visual extinction,  $A_V \approx 30$  magnitudes<sup>7</sup>, caused by dust and gas along the line of sight. Sgr A\* has not been detected in the infrared owing to its faintness and to the bright infrared background from stars and clouds of dust<sup>8</sup>. We thus need to detect X-rays from Sgr A\* in order to constrain the spectrum at energies above the radio-to-submillimetre band and to test whether gas is accreting onto a supermassive black hole (see above).

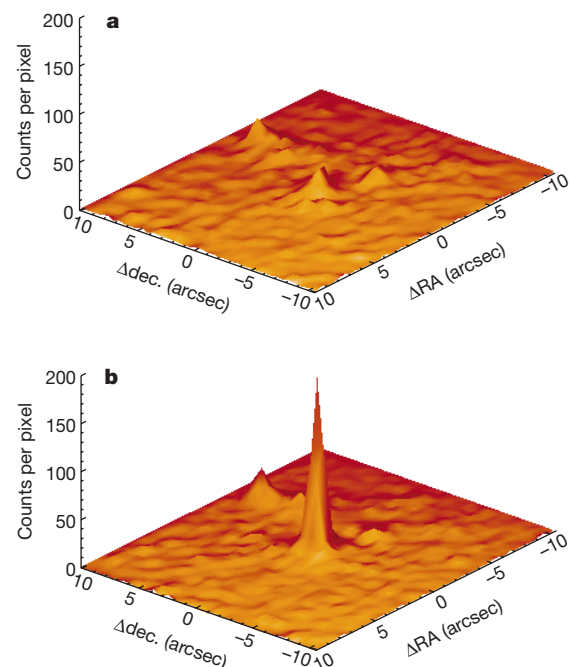
We first observed the Galactic Centre on 21 September 1999 with the imaging array of the Advanced CCD (charge-coupled device) Imaging Spectrometer (ACIS-I) aboard the Chandra X-ray Observatory<sup>9</sup> and discovered an X-ray source coincident within  $0.35'' \pm 0.26''$  ( $1\sigma$ ) of the radio source<sup>6</sup>. The luminosity in 1999 was very weak— $L_X \approx 2 \times 10^{33}$  erg s<sup>-1</sup> in the 2–10 keV band—after correction for the inferred neutral hydrogen absorption column  $N_H \approx 1 \times 10^{23}$  cm<sup>-2</sup>. This is far fainter than previous X-ray observatories could detect<sup>6</sup>.

We observed the Galactic Centre a second time with Chandra ACIS-I from 26 October 2000 22:29 until 27 October 2000 08:19 (UT), during which time we saw a source at the position of Sgr A\* brighten dramatically for a period of approximately 10,000 s (10 ks). Figure 1 shows surface plots for both epochs of the 2–8 keV counts integrated over time from a  $20'' \times 20''$  region centred on the radio position of Sgr A\*. The modest peak of the integrated counts at Sgr A\* in 1999 increased by a factor of about 7 in 2000, despite the 12% shorter exposure. The peak integrated counts of the fainter features in the field show no evidence of strong variability, demonstrating that the flaring at Sgr A\* is intrinsic to the source.

Figure 2 shows light curves of the photon arrival times from the

direction of Sgr A\* during the observation in 2000. Figure 2a and b shows hard-band (4.5–8 keV) and soft-band (2–4.5 keV) light curves constructed from counts within an angular radius of  $1.5''$  of Sgr A\*. Both bands exhibit roughly constant, low-level emission for the first 14 ks or so, followed by a 6-ks period of enhanced emission beginning with a 500-s event ( $4.4\sigma$  significance using 150-s bins). At 20 ks, flare(s) of large relative amplitude occur, lasting about 10 ks, and finally the emission drops back to the low state for the remaining 6 ks or so. About 26 ks into the observation, the hard-band light curve drops abruptly by a factor of 5 within a span of around 600 s and then partially recovers within a period of 1.2 ks. The soft-band light curve shows a similar feature, but it appears to lag the hard-band event by a few hundred seconds and is less sharply defined.

The band-ratio time series in Fig. 2c, defined as the ratio of hard-band counts to soft-band counts, suggests that the spectrum 'hardened' (that is, became flatter and extended more strongly to higher energies) during the flare. The difference between the band ratio measured at the peak of the flare and the average of the band ratios during the quiescent periods at the beginning and end of the observation is  $0.63 \pm 0.21$  (that is,  $3\sigma$ ). The peak-flare band ratio in Fig. 2c is affected to some extent by the effects of pile-up (see Fig. 2 legend), which would tend to harden the spectrum; however, the band ratios in Fig. 2d, which were computed using the non-piled-up data extracted from the wings of the point spread function, also show evidence for spectral hardening with  $2.7\sigma$  significance. The sizes of dust-scattering haloes in the Galactic Centre are typically



**Figure 1** Surface plots of the 2–8 keV counts within a  $20'' \times 20''$  field centred on Sgr A\* at two epochs. The data were taken with Chandra ACIS-I on (a) 21 September 1999 and (b) 26–27 October 2000. The effective exposure times were 40.3 ks and 35.4 ks, respectively. The spatial resolution is  $0.5''$  per pixel. An angle of  $1''$  on the sky subtends a projected distance of about 0.04 pc at the galactocentric distance of 8.0 kpc (ref. 30). The peak integrated counts per pixel at the position of Sgr A\* increased by a factor of 7 from the first epoch to the second, despite the slightly smaller exposure time ( $\sim 12\%$ ) in the second epoch. The low-level peak a few arcseconds to the southwest of Sgr A\* is the infrared source IRS 13, and the ridge of emission to the northwest is from a string of unresolved point sources. The fainter features in the field are reasonably consistent between the two epochs, considering the limited Poisson statistics and the fact that these stellar sources may themselves be variable; this consistency shows that the strong variations at Sgr A\* are intrinsic to the source.

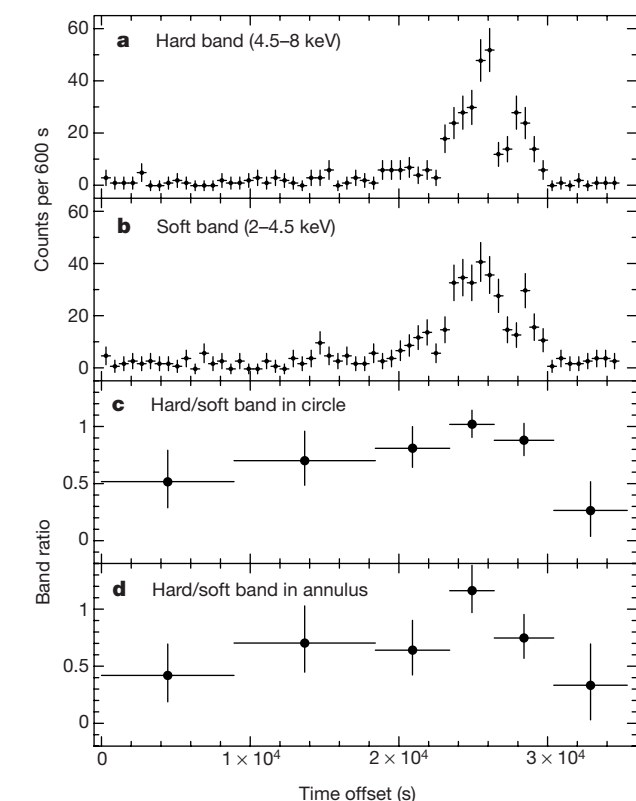


greater than  $1'$  (ref. 10), so dust-scattered X-rays from the source contribute a negligible fraction of the emission within the source extraction region that we used; hence dust-scattered X-rays cannot account for the spectral variations. We therefore conclude that the spectral hardening during the flare is likely to be real.

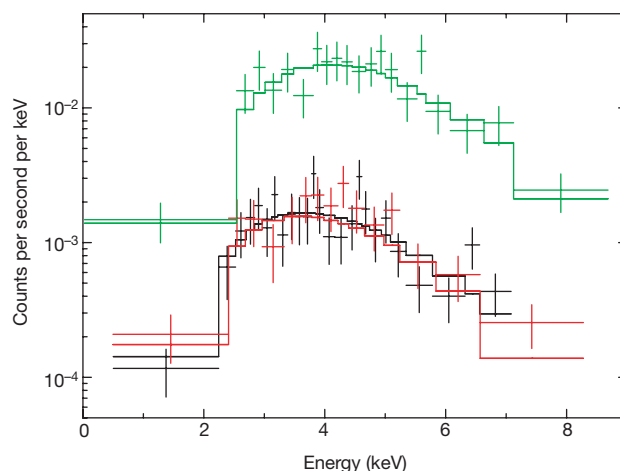
The quiescent-state spectra in 1999 and 2000 and the peak flaring-state spectrum in 2000 are shown in Fig. 3. We fitted each spectrum individually using a single power-law model with corrections for the effects of photoelectric absorption and dust scattering<sup>10</sup>. The best-fit values and 90% confidence limits for the parameters of each fit are presented in the first three lines of Table 1. The column densities for the three spectra are consistent, within the uncertainties, as are the photon indices of both quiescent-state spectra. Next, we fit a double power-law model to the three spectra simultaneously, using a single photon index for both quiescent spectra, a second photon index for the flaring spectrum, and a single column density for all three spectra. The best-fit models for each spectrum from the simultaneous fits are shown as solid lines in Fig. 3; the parameter values are given in the last line of Table 1. Using these values, we derive an absorption-corrected 2–10 keV luminosity of  $L_X = (2.2^{+0.4}_{-0.3}) \times 10^{33} \text{ erg s}^{-1}$  for the quiescent-state emission and  $L_X = (1.0 \pm 0.1) \times 10^{35} \text{ erg s}^{-1}$  for the peak of the flaring-state emission, or around 45 times the quiescent-state luminosity. We note that previous X-ray observatories did not have the sensitivity to detect such a short-duration, low-luminosity flare in the Galactic Centre<sup>6</sup>. The best-fit photon index  $\Gamma_f = 1.3^{+0.5}_{-0.6}$  ( $N(E) \propto E^{-\Gamma}$ ) for the flaring-state spectrum is slightly flatter than, but consistent with, systems thought to contain supermassive black holes<sup>11</sup>. Here  $E$  is the energy of a photon in keV,  $N(E)$  is the number of photons with energies in the differential interval  $E$  to  $E + dE$ , and  $\Gamma$  is the power-law index of the photon number distribution.

If we view the outburst as a single event, the rise/fall timescales of a few hundred seconds and the 10-ks duration are consistent with the light-crossing and dynamical timescales for the inner part (at less than 10 Schwarzschild radii;  $R_s \equiv 2GM/c^2$ ) of the accretion flow around a black hole of  $2.6 \times 10^6$  solar masses; here  $R_s$  is the radius of the black-hole event horizon (that is, the boundary at which the escape velocity equals the speed of light,  $c$ ),  $G$  is the gravitational constant and  $M$  is the mass of the black hole. Although we cannot strictly rule out an unrelated contaminating source as the origin of the flare (for example, an X-ray binary, for which little is known about such short-timescale, low-luminosity events as we have detected; W. Lewin, personal communication), this explanation seems unlikely, as the characteristic angular scales of the young and old stellar clusters around Sgr A\* are 5–20" (ref. 7), whereas the

osity of  $L_X = (2.2^{+0.4}_{-0.3}) \times 10^{33} \text{ erg s}^{-1}$  for the quiescent-state emission and  $L_X = (1.0 \pm 0.1) \times 10^{35} \text{ erg s}^{-1}$  for the peak of the flaring-state emission, or around 45 times the quiescent-state luminosity. We note that previous X-ray observatories did not have the sensitivity to detect such a short-duration, low-luminosity flare in the Galactic Centre<sup>6</sup>. The best-fit photon index  $\Gamma_f = 1.3^{+0.5}_{-0.6}$  ( $N(E) \propto E^{-\Gamma}$ ) for the flaring-state spectrum is slightly flatter than, but consistent with, systems thought to contain supermassive black holes<sup>11</sup>. Here  $E$  is the energy of a photon in keV,  $N(E)$  is the number of photons with energies in the differential interval  $E$  to  $E + dE$ , and  $\Gamma$  is the power-law index of the photon number distribution.



**Figure 2** Light curves of the photon arrival times and band ratios from the direction of Sgr A\* on 26–27 October 2000. **a**, Hard-band (4.5–8 keV) counts. **b**, Soft-band (2–4.5 keV) counts. **c**, Hard/soft band ratio within a circle of radius 1.5". **d**, Hard/soft band ratio within an annulus of inner and outer radii 0.5" and 2.5", respectively. The x axis shows the time offset from the start of the observation at 26 October 2000 22:29 (UT). The data are shown with  $1\sigma$  error bars. The single 2.8-hour period of flaring activity which we have detected so far during a total of 21 hours of observations yields a (poorly determined) duty cycle of approximately 1/8. Our continuing observations of this source with Chandra will permit us to refine this value. During the quiescent intervals at the beginning and end of this observation, the mean count rate in the 2–8 keV band was  $(6.4 \pm 0.6) \times 10^{-3} \text{ counts s}^{-1}$ , consistent with the count rate we measured in 1999, which was  $(5.4 \pm 0.4) \times 10^{-3} \text{ counts s}^{-1}$ . The detected count rate at the peak of the flare was  $0.16 \pm 0.01 \text{ counts s}^{-1}$  within an extraction circle of radius 1.5", but this is about 30% less than the true incident count rate owing to pile-up of X-rays in the detector during the 3.2-s integration time for each CCD (charge-coupled device) read out.



**Figure 3** X-ray spectra of the Chandra source at the position of Sgr A\*. The data are shown as crosses with vertical bars indicating the  $1\sigma$  errors in the count rate and horizontal bars the energy range of each bin. The events have been grouped to yield 10 counts per bin. The counts in the 1999 (black) and 2000 (red) quiescent-state spectra were extracted using a source radius of 1.5". A non-piled-up, peak flaring-state spectrum (green) was extracted from the wings of the point spread function, using an annulus with inner and outer radii of 0.5" and 2.5", during the time interval 23.7–26.3 ks after the start of the 2000 observation. The solid lines are the best-fit models for each spectrum, obtained by fitting an absorbed, dust-scattered, double power-law model to the three spectra simultaneously. The best-fit values and 90% confidence intervals for the model parameters are given in the last line of Table 1. The spectra are well-fitted using a single photon index for both quiescent spectra, a second photon index for the flaring spectrum, and a single column density for all three spectra. The column density from the simultaneous fits corresponds to a visual extinction  $A_V = 29.6^{+5.0}_{-6.1}$  magnitudes, which agrees well with infrared-derived estimates of  $A_V \approx 30$  magnitudes<sup>7</sup>; we thus find no evidence in our X-ray data of excess gas and dust localized around the supermassive black hole. This places an important constraint on the maximum contribution to the infrared spectrum of Sgr A\* produced by local dust reprocessing of higher energy photons from the accretion flow. The spectral models used in ref. 6 did not account for dust scattering; hence a higher column density was needed to reproduce the low-energy cut-off via photoelectric absorption alone. We note that there is no sign of an iron K $\alpha$  emission line in the flaring-state spectrum.

flaring source lies within  $1/3''$  of the radio position. These clusters contain up to a million solar masses of stars and stellar remnants<sup>2</sup>; hence it is rather improbable that there would be only one very unusual stellar X-ray source in the image and that it would be fortuitously superposed on Sgr A\*. Furthermore, it is not clear that X-ray binaries can be easily formed or long endure near Sgr A\*, given the high velocity dispersion and high spatial density of the stars in its deep gravitational potential well<sup>6,12</sup>.

Strong, variable X-ray emission is a characteristic property of active galactic nuclei; factors of 2–3 variations on timescales ranging from minutes to years are typical for radio-quiet active galactic nuclei<sup>11</sup>. Moderate- to high-luminosity active galactic nuclei (that is, Seyfert galaxies and quasars) show a general trend of increasing variability with decreasing luminosity<sup>13</sup>. However, this trend does not extend to low-luminosity active galactic nuclei (LLAGN), which show little or no significant variability on timescales less than a day<sup>14</sup>. Assuming the X-ray flare is from Sgr A\*, it is remarkable that this source—generally thought to be the nearest and least luminous example of accretion onto a central supermassive black hole—has shown a factor of 45 variation that is an order of magnitude more rapid than the fastest observed variation of similar relative amplitude by a radio-quiet active galactic nucleus of any luminosity class<sup>15</sup>. We note that flares of similar luminosity would be undetectable by Chandra in the nucleus of even the nearest spiral galaxy, M31. LLAGN emit  $L_X \geq 10^{38} \text{ erg s}^{-1}$  (ref. 14), so it should be kept in mind that the astrophysics of accretion onto even the LLAGN may differ substantially from that of Sgr A\*. This makes Sgr A\* a useful source for testing the theory of accretion onto supermassive black holes in galactic nuclei.

The faintness of Sgr A\* at all wavelengths requires that the supermassive black hole be in an extremely quiet phase, either because the accretion rate is very low, or because the accretion flow is radiatively inefficient, or both<sup>2</sup>. A variety of theoretical scenarios, usually based on advective accretion models<sup>16–19</sup>, jet-disk models<sup>20</sup>, or Bondi–Hoyle models<sup>21,22</sup>, have developed this idea. An important prediction of the advective accretion models is that the X-ray spectrum in the Chandra energy band should be dominated by thermal bremsstrahlung emission from hot gas in the outer regions of the accretion flow ( $R \geq 10^3 R_S$ ), but a region this large could not produce the rapid, large-relative-amplitude variations we have seen. Thus, the properties of the X-ray flare are inconsistent with the advective accretion flow models. The low luminosity and short timescales of this event are also inconsistent with tidal disruption of a star by a central supermassive black hole<sup>23</sup>.

In all models, the radio-to-submillimetre spectrum of Sgr A\* is cyclo-synchrotron emission from a combination of sub-relativistic and relativistic electrons (and perhaps positrons) spiralling around magnetic field lines either in a jet or in a static region within the inner  $10 R_S$  of the accretion flow. The electron Lorentz factor inferred from the radio spectrum of Sgr A\* is  $\gamma_e \approx 10^2$ . If the X-ray flare were produced via direct synchrotron emission, then the emitting electrons would need  $\gamma_e \geq 10^5$ . For the 10–100 G magnetic field strengths predicted by the models, the cooling time of the particles would be  $\sim 1$ –100 s. Thus, the approximately 10-ks

duration of the flare would require repeated injection of energy to the electrons. On the other hand, if the X-rays were produced via up-scattering of the submillimetre photons off the relativistic electrons, a process called synchrotron self-Comptonization (SSC), then  $\gamma_e \approx 10^2$ – $10^3$  would be required, and the cooling time would be of the order of hours, which is consistent with the duration of the flare. The rapid turn-off of the X-ray emission might then be attributed to the dilution of both photon and electron densities in an expanding plasma.

The X-ray spectra of radio-quiet quasars and active galactic nuclei are thought to be produced by thermal Comptonization of infrared-to-ultraviolet seed photons from a cold, optically thick, geometrically thin accretion disk by hot electrons in a patchy corona above the disk<sup>13,24</sup>. The X-ray spectra of these sources generally ‘soften’ (that is, become steeper and extend less strongly to higher energies) as they brighten<sup>15</sup>. In contrast, the extremely low luminosity of Sgr A\* precludes the presence of a standard, optically thick accretion disk<sup>5</sup>; hence, the dominant source of seed photons would be the millimetre-to-submillimetre synchrotron photons.

The energy released by an instability in the mass accretion rate or by a magnetic reconnection event near the black hole would shock-accelerate the electrons, causing the synchrotron spectrum to intensify and to extend farther into the submillimetre band. Consequently, the Compton up-scattered X-ray emission would harden as the X-ray intensity increased, exactly as observed. We note that the millimetre-band spectrum of Sgr A\* has been observed to harden during one three-week flare<sup>25</sup> and one three-day flare<sup>26</sup>, as would be required by the current SSC models for Sgr A\* (refs 20, 22).

To test the SSC models, we measured the flux density of Sgr A\* at a wavelength of 3 mm with the Millimeter Array at the Owens Valley Radio Observatory, simultaneous with part of the 2000 Chandra measurement. Unfortunately, the available observing window (20:10–02:30 UT) preceded the X-ray flare (04:03–06:50 UT) by a few hours. The observed flux density of Sgr A\* was  $2.05 \pm 0.3 \text{ Jy}$ , consistent with previously reported measures<sup>27,28</sup>. Recently, a 106-day quasi-periodicity has been reported in the centimetre band from an analysis of 20 years of data taken with the Very Large Array (VLA)<sup>29</sup>. A weekly VLA monitoring program detected a 30% increase in the radio flux density of Sgr A\* beginning around 24 October 2000 and peaking on 5 November 2000. This increase was seen at 2 cm, 1.3 cm, and 7 mm (R. McGary, J.-H. Zhao, W. M. Goss and G. C. Bower, personal communication). The timing of the X-ray flare and the rise in the radio flux density of Sgr A\* suggests that there is a connection between the two events, providing additional indirect support for the association of the X-ray flare with Sgr A\* and further strengthening the case that it was produced via either the SSC or direct synchrotron processes. Definitive evidence for these ideas will require detection of correlated variations in the radio-to-submillimetre and X-ray wavebands through future coordinated monitoring projects. □

Received 27 April; accepted 2 August 2001.

**Table 1 Spectral fits**

| Spectrum       | $N_H$               | $\Gamma_q$          | $\Gamma_f$          | $\chi^2/\nu$ |
|----------------|---------------------|---------------------|---------------------|--------------|
| 1999 quiescent | $5.8^{+1.3}_{-1.4}$ | $2.5^{+0.8}_{-0.7}$ | ...                 | 19/22        |
| 2000 quiescent | $5.0^{+1.9}_{-2.3}$ | $1.8^{+0.7}_{-0.9}$ | ...                 | 7.6/12       |
| 2000 flaring   | $4.6^{+2.0}_{-1.6}$ | ...                 | $1.0^{+0.8}_{-0.7}$ | 12/17        |
| All            | $5.3^{+0.9}_{-1.1}$ | $2.2^{+0.5}_{-0.7}$ | $1.3^{+0.5}_{-0.6}$ | 45/55        |

Best-fit parameter values and 90% confidence intervals for power-law models, corrected for photoelectric absorption and dust scattering<sup>10</sup>.  $N_H$  is the neutral hydrogen absorption column in units of  $10^{22} \text{ H atoms cm}^{-2}$ .  $\Gamma_q$  and  $\Gamma_f$  are the photon-number indices of the quiescent-state and peak flaring-state spectra ( $N(E) \propto E^{-\Gamma}$ ).  $\chi^2$  is the value of the fit statistic for the best-fit model, and  $\nu$  is the number of degrees of freedom in the fit. The parameter values for the spectrum marked ‘All’ were derived by fitting an absorbed, dust-scattered, double power-law model to the three spectra simultaneously (see text and Fig. 3).

1. Richstone, D. *et al.* Supermassive black holes and the evolution of galaxies. *Nature* **395** (suppl. on optical astronomy) A14–A19 (1998).
2. Genzel, R., Pichon, C., Eckart, A., Gerhard, O. E. & Ott, T. Stellar dynamics in the Galactic Centre: proper motions and anisotropy. *Mon. Not. R. Astron. Soc.* **317**, 348–374 (2000).
3. Ghez, A. M., Morris, M., Becklin, E. E., Tanner, A. & Kremenek, T. The accelerations of stars orbiting the Milky Way’s central black hole. *Nature* **407**, 349–351 (2000).
4. Lynden-Bell, D. & Rees, M. J. On quasars, dust and the Galactic Centre. *Mon. Not. R. Astron. Soc.* **152**, 461–475 (1971).
5. Melia, F. & Falcke, H. The supermassive black hole at the Galactic Center. *Annu. Rev. Astron. Astrophys.* **39**, 309–352 (2001).
6. Baganoff, F. K. *et al.* Chandra X-ray spectroscopic imaging of Sgr A\* and the central parsec of the Galaxy. *Astrophys. J.* (submitted); also preprint astro-ph/0102151 at (xxx.lanl.gov) (2001).
7. Morris, M. & Serabyn, E. The galactic center environment. *Annu. Rev. Astron. Astrophys.* **34**, 645–702 (1996).
8. Menten, K. M., Reid, M. J., Eckart, A. & Genzel, R. The position of Sagittarius A\*: accurate alignment of the radio and infrared reference frames at the Galactic Center. *Astrophys. J.* **475**, L111–L114 (1997).
9. Weisskopf, M. C., O’Dell, S. L. & van Speybroeck, L. P. Advanced X-Ray Astrophysics Facility (AXAF). *Proc. SPIE* **2805**, 2–7 (1996).

10. Predehl, P. & Schmitt, J. H. M. M. X-raying the interstellar medium: ROSAT observations of dust scattering halos. *Astron. Astrophys.* **293**, 889–905 (1995).
11. Mushotzky, R. F., Done, C. & Pounds, K. A. X-ray spectra and time variability of active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **31**, 717–761 (1993).
12. Davies, M. B., Blackwell, R., Bailey, V. C. & Sigurdsson, S. The destructive effects of binary encounters on red giants in the Galactic Centre. *Mon. Not. R. Astron. Soc.* **301**, 745–753 (1998).
13. Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J. & Yaqoob, T. ASCA observations of Seyfert 1 galaxies. I. Data analysis, imaging, and timing. *Astrophys. J.* **476**, 70–82 (1997).
14. Ptak, A., Yaqoob, T., Mushotzky, R., Serlemitsos, P. & Griffiths, R. X-ray variability as a probe of advection-dominated accretion in low-luminosity active galactic nuclei. *Astrophys. J.* **501**, L37–L40 (1998).
15. Ulrich, M.-H., Maraschi, L. & Urry, C. M. Variability of active galactic nuclei. *Annu. Rev. Astron. Astrophys.* **35**, 445–502 (1997).
16. Narayan, R., Mahadevan, R., Grindlay, J. E., Popham, R. G. & Gammie, C. Advection-dominated accretion model of Sagittarius A\*: evidence for a black hole at the Galactic center. *Astrophys. J.* **492**, 554–568 (1998).
17. Quataert, E. & Narayan, R. Spectral models of advection-dominated accretion flows with winds. *Astrophys. J.* **520**, 298–315 (1999).
18. Ball, G. H., Narayan, R. & Quataert, E. Spectral models of convection-dominated accretion flows. *Astrophys. J.* **552**, 221–226 (2001).
19. Blandford, R. D. & Begelman, M. C. On the fate of gas accreting at a low rate on to a black hole. *Mon. Not. R. Astron. Soc.* **303**, L1–L5 (1999).
20. Falcke, H. & Markoff, S. The jet model for Sgr A\*: radio and X-ray spectrum. *Astron. Astrophys.* **362**, 113–118 (2000).
21. Melia, F. An accreting black hole model for Sagittarius A\*. II: A detailed study. *Astrophys. J.* **426**, 577–585 (1994).
22. Melia, F., Liu, S. & Coker, R. Polarized millimeter and submillimeter emission from Sagittarius A\* at the Galactic center. *Astrophys. J.* **545**, L117–L120 (2000).
23. Rees, M. J. Tidal disruption of stars by black holes of  $10^6$ – $10^8$  solar masses in nearby galaxies. *Nature* **333**, 523–528 (1988).
24. Haardt, F., Maraschi, L. & Ghisellini, G. X-ray variability and correlations in the two-phase disk-corona model for Seyfert galaxies. *Astrophys. J.* **476**, 620–631 (1997).
25. Tsuboi, M., Miyazaki, A. & Tsutsumi, T. In *The Central Parsecs of the Galaxy* (ed. Falcke, H. et al.) Vol. 186, 105–112 (ASP Conf. Ser., Astronomical Society of the Pacific, San Francisco, 1999).
26. Wright, M. C. H. & Backer, D. C. Flux density of Sagittarius A at  $\lambda = 3$  millimeters. *Astrophys. J.* **417**, 560–564 (1993).
27. Serabyn, E. et al. High frequency measurements of the spectrum of SGR A\*. *Astrophys. J.* **490**, L77–L81 (1997).
28. Falcke, H. et al. The simultaneous spectrum of Sagittarius A\* from 20 centimeters to 1 millimeter and the nature of the millimeter excess. *Astrophys. J.* **499**, 731–734 (1998).
29. Zhao, J.-H., Bower, G. C. & Goss, W. M. Radio variability of Sagittarius A\*—a 106 day cycle. *Astrophys. J.* **547**, L29–L32 (2001).
30. Reid, M. J. The distance to the center of the Galaxy. *Annu. Rev. Astron. Astrophys.* **31**, 345–372 (1993).

## Acknowledgements

We thank M. Begelman for useful comments. This work has been supported by a grant from NASA.

Correspondence and requests for materials should be addressed to F.K.B. (e-mail: fkb@space.mit.edu).

## How ‘spin ice’ freezes

J. Snyder\*, J. S. Slusky†, R. J. Cava† & P. Schiffer\*

\* Department of Physics and Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA

† Department of Chemistry and Princeton Materials Institute, Princeton University, Princeton, New Jersey 08540, USA

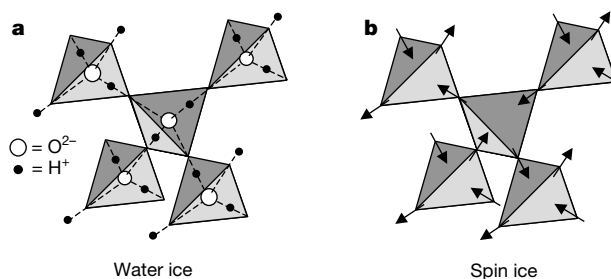
The large degeneracy of states resulting from the geometrical frustration of competing interactions is an essential ingredient of important problems in fields as diverse as magnetism<sup>1</sup>, protein folding<sup>2</sup> and neural networks<sup>3</sup>. As first explained by Pauling<sup>4</sup>, geometrical frustration of proton positions is also responsible for the unusual low-temperature thermodynamics of ice and its measured ‘ground state’ entropy<sup>5</sup>. Recent work has shown that the geometrical frustration of ice is mimicked by  $\text{Dy}_2\text{Ti}_2\text{O}_7$ , a site-ordered magnetic material in which the spins reside on a lattice of corner-sharing tetrahedra where they form an unusual magnetic ground state known as ‘spin ice’<sup>6–13</sup>. Here we identify a cooperative spin-freezing transition leading to the spin-ice ground state in

$\text{Dy}_2\text{Ti}_2\text{O}_7$ . This transition is associated with a very narrow range of relaxation times, and represents a new form of spin-freezing. The dynamics are analogous to those associated with the freezing of protons in ice, and they provide a means through which to study glass-like behaviour and the consequences of frustration in the limit of low disorder.

In the ground state of ordinary hexagonal ice ( $\text{I}_h$ ), neighbouring oxygen ions are connected through a hydrogen bond (Fig. 1a), and each hydrogen ion (proton) is situated closer to one or the other of its oxygen nearest-neighbours. This situation was codified by Pauling<sup>4</sup> as the ‘ice rules’: that there should be a proton between each oxygen pair and that it should be localized near one or the other oxygen site. As each oxygen ion has four oxygen nearest-neighbours, the ice rules require that two of the neighbouring protons will be close to, and two will be far from, any particular oxygen ion. There is a six-fold degeneracy of protonic states that obey this rule for any given oxygen site, and a macroscopic degeneracy of allowed states in an extended crystal. Although the proton positions are dynamic at high temperatures owing to mobile protonic point defects (for example, two or zero interstitial protons between oxygen sites), the dynamics slow substantially with decreasing temperature, and the protons effectively freeze into a disordered configuration below  $T \approx 100$  K (refs 14, 15). The metastable nature of this frozen state leads to the observation of a ‘practical’ ground-state entropy, associated with the random configuration of the protons as  $T \rightarrow 0$ , which is in almost exact agreement with that expected from the ice rules<sup>15,16</sup>.

The statistical mechanics of the ice rules may be mimicked by certain rare-earth magnets with the pyrochlore structure<sup>6</sup>. In such a spin-ice material, the magnetic rare-earth ions are situated on a lattice of corner-sharing tetrahedra and their spins are constrained by crystal field interactions<sup>11,17</sup> to point either directly towards or directly away from the centres of the tetrahedra (Fig. 1b). Dipole and ferromagnetic exchange interactions between the spins require that, on each tetrahedron, two spins point inward and two point outward. This situation exactly mirrors the frustration associated with the ice rules, and leads to the same large degeneracy of states<sup>10–13</sup>. The spin-ice state has been realized experimentally<sup>9,11</sup> in  $\text{Dy}_2\text{Ti}_2\text{O}_7$ , in which the magnetic specific heat integrated from zero temperature yields a measured ground-state entropy in exact agreement with the theoretical prediction for the ice rules and experimental results for ice<sup>5,15,16</sup>.

In order to examine the development of the magnetic ground state of spin ice, we study the magnetization ( $M$ ) and the resultant d.c. susceptibility ( $\chi_{dc} = dM/dH$ , where  $H$  is the applied magnetic field) as well as the real and imaginary parts ( $\chi'$  and  $\chi''$ ) of the a.c. susceptibility ( $\chi_{ac}$ ) of polycrystalline  $\text{Dy}_2\text{Ti}_2\text{O}_7$ . The calorimetric measurements suggest spin-freezing into a non-equilibrium state<sup>9,10</sup>, but we find that  $\chi_{dc}$  increases monotonically with decreas-

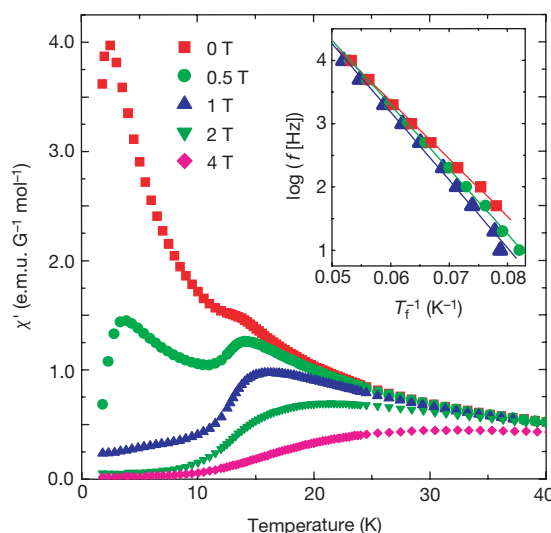


**Figure 1** Schematic representation of frustration in water ice and spin ice. **a**, In water ice, each hydrogen ion is close to one or the other of its two oxygen neighbours, and each oxygen must have two hydrogen ions closer to it than to its neighbouring oxygen ions. **b**, In spin ice, the spins point either directly toward or away from the centres of the tetrahedra, and each tetrahedron is constrained to have two spins pointing in and two pointing out.

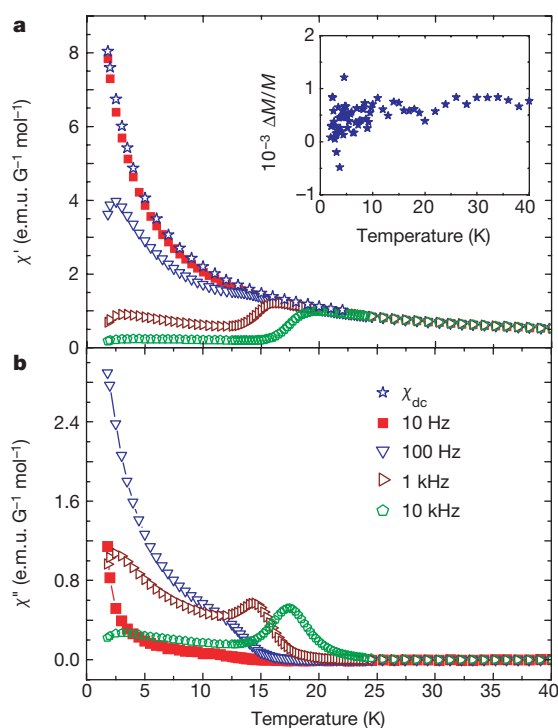


ing temperature as expected for a paramagnetic system with no spin-freezing (Fig. 2). Although  $\chi'(T)$  is virtually identical to  $\chi_{dc}$  at our lowest frequency, at higher frequencies we observe that  $\chi'(T)$  decreases sharply at  $T \approx 15$  K, deviating well below  $\chi_{dc}$ . This sharp decrease leads to a local maximum in  $\chi'(T)$  correlated with a sharp rise in  $\chi''(T)$  at a 'freezing' temperature,  $T_f$ , which increases with frequency. This feature is a common signature of a glass transition in both structural glasses and spin glasses, and it is also observed in the dielectric permittivity of ice (which directly probes the local protonic motion)<sup>14</sup>. The drop in  $\chi'(T)$  indicates that the spins' dynamic response is slowed such that they cannot respond to the time-varying magnetic field for  $T < T_f$ . This implies that the system is out of equilibrium on the timescale of the measurement, and thus that the observed properties depend on that timescale—a classic signature of glassiness.

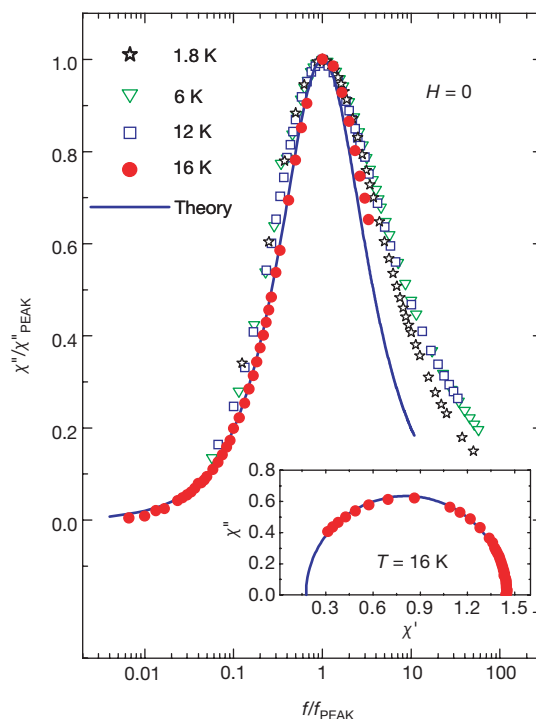
As we observe a glass-like freezing of the spin degrees of freedom, the simplest interpretation would be that this is a spin-glass transition. Such a transition in a pure material (with an expected randomness of less than 1% in either structural or site disorder) would be unusual, but is not unprecedented in geometrically frustrated magnets<sup>18–21</sup>. As is the case for many glass-like systems, the frequency ( $f$ ) dependence of  $T_f$  can be fitted to an Arrhenius relation,  $f = f_0 \exp(-E_a/k_B T_f)$ , with physically reasonable values of  $f_0 = 720$  MHz and  $E_a/k_B = 210$  K at  $H = 0$  (Fig. 3 inset). The frequency dependence can be quantified as  $p = \Delta T_f / T_f \Delta(\log f_0) \approx 0.18$ , which is much stronger than is typical for spin glasses (where  $p \approx 0.01$ ; ref. 22) and is more typical of spin-freezing in dilute dipole glasses<sup>23</sup> or superparamagnetic blocking (both of which are qualitatively rather different from this dense frustrated system). This strong frequency dependence, also seen in the dielectric permittivity of ice<sup>14</sup>, suggests an explanation for the absence of a freezing signature in  $\chi_{dc}$  (Fig. 2 inset) and the much lower temperature of the observed development of spin correlations in the heat capacity data of



**Figure 3** Real part of the a.c. magnetic susceptibility of  $\text{Dy}_2\text{Ti}_2\text{O}_7$  versus temperature for several different fields at 100 Hz. Inset, measurement frequency versus the inverse of the freezing temperature with fits to an Arrhenius law in magnetic fields of 0, 5,000 and  $10^4$  G. Note from both plots that the application of a magnetic field raises the freezing temperature, opposite to the effect on a spin glass.



**Figure 2** Temperature dependence of the magnetic susceptibility in the absence of a d.c. magnetic field. **a**, The d.c. magnetic susceptibility ( $\chi_{dc}$ ) and the real part of the a.c. magnetic susceptibility ( $\chi'$ ). **b**, The imaginary part of the a.c. magnetic susceptibility ( $\chi''$ ). Inset, the normalized difference between the magnetization at 100 G when the sample is cooled in a field and when it is cooled in the absence of the field. The deviation from zero is within the margin of reproducibility of the measurement.



**Figure 4** The imaginary part of the magnetic susceptibility scaled to peak amplitude and frequency for  $\text{Dy}_2\text{Ti}_2\text{O}_7$  at several temperatures and in the absence of a d.c. magnetic field. Also shown is the peak expected from the Casimir-du Pré relation assuming a single characteristic relaxation time for the system. Inset, Cole-Cole plot of the susceptibility data at different frequencies for  $T = 16$  K. The semicircular character of the plot demonstrates the narrow distribution of relaxation times around a characteristic relaxation time of about 5 ms. The solid line in the inset is a fit to a semicircle which yields values for the isothermal and adiabatic susceptibilities of 0.18 and  $1.4 \text{ e.m.u. G}^{-1} \text{ mol}^{-1}$ , respectively.

ref. 9—as those two measurements operate on timescales decades longer than that of our minimum frequency for  $\chi_{ac}$ .

One difference between the observed spin-ice freezing and glassy freezing in other magnets is the effect of magnetic field. Whereas a magnetic field suppresses the freezing temperature in spin glasses and in superparamagnets<sup>19,22,24</sup>, we see in Fig. 3 that the application of a field in fact enhances  $T_f$  and suppresses  $\chi''(T)$  strongly for all temperatures below  $T_f$ , even at our lowest measurement frequency. Correspondingly, both fitting parameters  $E_a$  and  $f_0$  increase with field, reflecting increasing barriers to relaxation in an applied field.

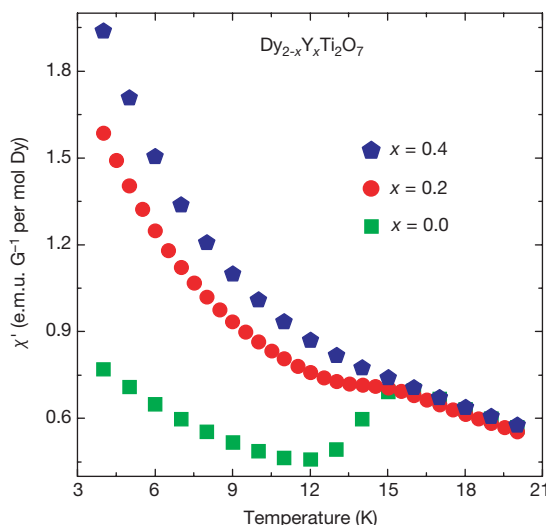
A more striking difference between spin-ice freezing and the glassy freezing observed in other magnets is found in the frequency distribution of the spin relaxation processes. In Fig. 4 we plot the frequency dependence of  $\chi''$  normalized to its maximum value. Such data are typically analysed in terms of a Casimir-du Pré relation<sup>25</sup> which predicts, for a single relaxation time  $\tau$ , that  $\chi''(f) = f\tau[(\chi_T - \chi_S)/(1 + f^2\tau^2)]$  where  $\chi_T$  is the isothermal susceptibility in the limit of low frequency and  $\chi_S$  is the adiabatic susceptibility in the limit of high frequency. We see that our data fit such a form at  $T = 16$  K, and the quality of the fit is confirmed by a Cole-Cole (Argand) plot of  $\chi'$  versus  $\chi''$  (Fig. 4 inset)—we find a close approximation of the theoretically expected semicircle corresponding to a single relaxation mode<sup>22</sup>. Furthermore, the distribution of modes is broadened only slightly by lowering the temperature: the full-width at half-maximum reaches only about 1.6 decades in frequency—close to the theoretical expectation of 1.14 decades for a single relaxation mode. This situation is in sharp contrast to that in other dense magnetic systems exhibiting glass-like behaviour, in which the relaxation times typically span several decades<sup>26–28</sup>.

The narrow range of relaxation times in  $\text{Dy}_2\text{Ti}_2\text{O}_7$  is analogous to the observed dynamic freezing of protons in ice, where the frequency-dependent dielectric permittivity can be described by a single relaxation time associated with motion of protonic point defects. In  $\text{Dy}_2\text{Ti}_2\text{O}_7$  the Dy spins are localized and highly uniaxial, and the mode of relaxation is presumably through single-spin reorientations between the axial directions. Given that the dipole interaction energy between the spins is of order 2 K (ref. 13) and the spin–spin exchange energy is of order 1 K (refs 8,9,12), the barrier to spin-flips in our Arrhenius-law fit must be associated with the crystal field splitting ( $\sim 300$  K; ref. 11) from which the spin anisotropy originates. What we are apparently observing is the freezing of defects in the spin-ice lattice, in analogy to the freezing of

point defects in ice: at high temperatures the spins can flip uncorrelated with their neighbours, but at low temperatures the correlations between the spins (that is, the ice rules) prevent such flips. We have demonstrated the importance of spin–spin correlations to the freezing through studies of samples in which the dysprosium ions were diluted with non-magnetic yttrium. In such samples, the freezing transition is strongly suppressed even for Y concentrations of less than 20% (Fig. 5).

The fact that we observe a finite a.c. susceptibility below  $T_f$  indicates that there are still spin fluctuations within the spin-ice state that allow the system to respond to the varying magnetic field of the susceptibility measurement—as long as the timescale of that variation is not too short. Although such fluctuations are expected owing to the high degeneracy of spin states in this strongly frustrated system, they also should be suppressed at a sufficiently low temperature. Indeed, we do see a slight suppression of  $\chi'$  at the lowest temperatures and a narrowing of the peak in the Casimir-du Pré plot, consistent with such a suppression.

The spin-freezing in  $\text{Dy}_2\text{Ti}_2\text{O}_7$  represents a new form of spin-freezing. Rather than the dynamic freezing associated with relaxations over a wide range of length and time scales as in manifestly disordered systems, spin ice shows macroscopic behaviour that is analogous to glassiness (that is, freezing observed at different temperatures on different timescales), but predominantly associated with a single relaxation mode. This finding is consistent with  $\text{Dy}_2\text{Ti}_2\text{O}_7$  being a pure material: the frustration arises from the local geometry of the lattice and not the large-scale disorder typically associated with dynamic spin-freezing in magnetic systems. A very recent paper by Bramwell *et al.*<sup>29</sup> strongly suggested that the  $\text{Ho}_2\text{Ti}_2\text{O}_7$  ground state comprised a realization of the spin-ice model, in contrast to earlier conclusions that it did not<sup>11</sup>. Susceptibility data on this system are qualitatively consistent with our results, although the frequency scales appear to be different<sup>30</sup>. Although the spin-freezing we report here is similar to the protonic freezing in ice,  $\text{Dy}_2\text{Ti}_2\text{O}_7$  represents a much simpler system than ice, because the spin degrees of freedom are much more limited than those allowed by the complex dynamics of the protons<sup>5</sup>. We have observed this spin-freezing in readily accessible ranges of timescales and temperatures, which should facilitate future studies of the spin-ice ground state: we hope that such studies will shed new light on the nature of glassiness, particularly in the limit of low disorder.



**Figure 5** The real part of the a.c. magnetic susceptibility of  $\text{Dy}_{2-x}\text{Y}_x\text{Ti}_2\text{O}_7$  as a function of temperature in the absence of a d.c. magnetic field. These data were measured at 500 Hz. The dilution of the magnetic dysprosium sublattice with even a small fraction of

non-magnetic yttrium ions suppresses the freezing transition. This shows that the freezing is due to spin–spin interactions.

**Note added in proof:** After the present work was accepted for publication, an independent study of the magnetic susceptibility of  $\text{Dy}_2\text{Ti}_2\text{O}_7$  was published by Matsuhira *et al.*<sup>31</sup>, offering a somewhat different interpretation of the behaviour. □

## Methods

We studied polycrystalline samples of  $\text{Dy}_2\text{Ti}_2\text{O}_7$ , prepared from high-purity  $\text{Dy}_2\text{O}_3$  and  $\text{TiO}_2$  powders mixed in stoichiometric proportion and heated at 1,350 °C in air for a period of one week with intermediate grindings. The samples were shown to be single-phase pyrochlore type by powder X-ray diffraction. We have also studied a single crystal grown in a travelling floating zone mirror furnace, and obtain qualitatively equivalent data.

The d.c. susceptibility measurements were performed on a Quantum Design MPMS SQUID magnetometer. The a.c. susceptibility measurements were made with a Quantum Design PPMS system with the ACMS option. The data were independent of the amplitude of the excitation field between  $10^{-4}$  and  $10^{-3}$  T, and the frequency independence of the susceptometer was verified with a DyO calibration sample. As the Dy spins are highly anisotropic in  $\text{Dy}_2\text{Ti}_2\text{O}_7$ , these measurements probe the vast majority of the spins in our polycrystalline samples, with the exception of those whose easy axis is directed perpendicular to the applied field.

Received 14 May; accepted 20 July 2001.

- Ramirez, A. P. in *Handbook of Magnetic Materials* Vol. 13 (ed. Buschow, K. J. H.) 423–520 (Elsevier Science, Amsterdam, 2001).
- Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. Theory of protein folding: The energy landscape perspective. *Ann. Rev. Phys. Chem.* **48**, 545–600 (1997).
- Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl Acad. Sci.* **81**, 3088–3092 (1984).
- Pauling, L. *The Nature of the Chemical Bond* 301–304 (Cornell Univ. Press, Ithaca, New York, 1945).
- Petrenko, V. F. & Whitworth, R. W. *Physics of Ice* (Clarendon, Oxford, 1999).
- Harris, M. J., Bramwell, S. T., McMorrow, D. E., Zeiske, T. & Godfrey, K. W. Geometrical frustration in the ferromagnetic pyrochlore  $\text{Ho}_2\text{Ti}_2\text{O}_7$ . *Phys. Rev. Lett.* **79**, 2554–2557 (1997).
- Harris, M. J., Bramwell, S. T., Holdsworth, P. C. W. & Champion, J. D. M. Liquid-gas critical behaviour in a frustrated pyrochlore ferromagnet. *Phys. Rev. Lett.* **81**, 4496–4499 (1998).
- Bramwell, S. T. & Harris, M. J. Frustration in Ising-type spin models on the pyrochlore lattice. *J. Phys. Condens. Matter* **10**, L215–L220 (1998).
- Ramirez, A. P., Hayashi, A., Cava, R. J., Siddharthan, R. & Shastri, B. S. Zero-point entropy in 'spin ice'. *Nature* **399**, 333–335 (1999).
- Harris, M. J. Taking the frustration out of ice. *Nature* **399**, 311–312 (1999).
- Siddharthan, R. *et al.* Ising pyrochlore magnets: low-temperature properties, "ice rules," and beyond. *Phys. Rev. Lett.* **83**, 1854–1857 (1999).
- den Hertog, B. C., Melko, R. G. & Gingras, M. J. P. Long range order at low temperature in dipolar spin ice. Preprint cond-mat/0009225 at (<http://xxx.lanl.gov>) (2000).
- Gingras, M. J. P. & den Hertog, B. C. Origin of spin ice behavior in Ising pyrochlore magnets with long range dipole interactions: an insight from mean-field theory. Preprint cond-mat/0012275 at (<http://xxx.lanl.gov>) (2000).
- Kawada, S. Dielectric anisotropy in ice Ih. *J. Phys. Soc. Jpn* **44**, 1881–1886 (1978).
- Haida, O., Matsuo, T., Hiroshi, S. & Seki, S. Calorimetric study of the glassy state X. Enthalpy relaxation at the glass-transition temperature of hexagonal ice. *J. Chem. Thermodyn.* **6**, 815–825 (1974).
- Giaque, W. F. & Stout, J. W. The entropy of water and the third law of thermodynamics. The heat capacity of ice from 15K to 273K. *J. Am. Chem. Soc.* **58**, 1144–1150 (1936).
- Rosenkranz, S. *et al.* Crystal-field interaction in the pyrochlore magnet  $\text{Ho}_2\text{Ti}_2\text{O}_7$ . *J. Appl. Phys.* **87**, 5914–5916 (2000).
- Ramirez, A. P., Espinosa, G. P. & Cooper, A. S. Strong frustration and dilution enhanced order in a quasi-2D spin glass. *Phys. Rev. Lett.* **64**, 2070–2073 (1990).
- Gardner, J. S. Glassy statics and dynamics in the chemically ordered pyrochlore anti-ferromagnet  $\text{Y}_2\text{Mo}_2\text{O}_7$ . *Phys. Rev. Lett.* **83**, 211–215 (1999).
- Schiffer, P. *et al.* Frustration induced spin freezing in a site-ordered magnet—gadolinium gallium garnet. *Phys. Rev. Lett.* **74**, 2379–2382 (1995).
- Wills, A. S. Long range ordering and representational analysis of the jarosites. *Phys. Rev. B* **63**, 064430–1–064430-12 (2001).
- Mydosh, J. A. *Spin Glasses: An Experimental Introduction* (Taylor & Francis, London, 1993).
- Reich, D. H., Rosenbaum, T. F. & Aeppli, G. Glassy relaxation without freezing in a random dipolar-coupled Ising magnet. *Phys. Rev. Lett.* **59**, 1969–1972 (1987).
- Fiorani, D., Tholence, J. & Dorman, J. L. Magnetic properties of small ferromagnetic particles ( $\text{Fe-Al}_2\text{O}_3$  granular thin films): Comparison with spin glass properties. *J. Phys. C* **19**, 5495–5507 (1986).
- Casimir, H. B. G. & du Pré, F. K. Note on the thermodynamic interpretation of paramagnetic relaxation phenomena. *Physica* **5**, 507–511 (1938).
- Dirkmaat, A. J. *et al.* Frequency-dependence of the ac susceptibility in the random anisotropy system  $\text{Dy}(\text{P}_{1-x}\text{V}_x)_2\text{O}_4$ . *Phys. Rev. B* **36**, 352–359 (1987).
- Huser, D., Wenger, L. E., van Duynveldt, A. J. & Mydosh, J. A. Dynamical behavior of the susceptibility around the freezing temperature in (Eu,Sr)S. *Phys. Rev. B* **27**, 3100–3103 (1983).
- Dekker, C., Arts, A. F. M., de Wijn, H. W., van Duynveldt, A. J. & Mydosh, J. A. Activated dynamics in a two-dimensional Ising spin glass:  $\text{Rb}_2\text{Cu}_{1-x}\text{Co}_x\text{F}_4$ . *Phys. Rev. B* **40**, 243–251 (1989).
- Bramwell, S. T. *et al.* Spin correlations in  $\text{Ho}_2\text{Ti}_2\text{O}_7$ : A dipolar spin ice system. *Phys. Rev. Lett.* **87**, 047205-1–047205-4 (2001).
- Matsuhira, K., Hinatsu, Y., Tenya, K. & Sakakibara, T. Low temperature magnetic properties of frustrated pyrochlore ferromagnets  $\text{Ho}_2\text{Sn}_2\text{O}_7$  and  $\text{Ho}_2\text{Ti}_2\text{O}_7$ . *J. Phys. Condens. Matter* **12**, L649–L656 (2000).
- Matsuhira, K., Hinatsu, Y. & Sakakibara, T. Novel dynamical magnetic properties in the spin ice compound  $\text{Dy}_2\text{Ti}_2\text{O}_7$ . *J. Phys. Condens. Matter* **13**, L737 (2001).

## Acknowledgements

We thank D.A. Huse, T. Rosenbaum and J. Banavar for discussions. This work was supported by the Army Research Office.

Correspondence and requests for materials should be addressed to P.S. (e-mail: schiffer@phys.psu.edu).

# Fluidity of water confined to subnanometre films

Uri Raviv, Pierre Laurat\* & Jacob Klein\*

Weizmann Institute of Science, Rehovot 76100, Israel

The fluidity of water in confined geometries is relevant to processes ranging from tribology to protein folding, and its molecular mobility in pores and slits has been extensively studied using a variety of approaches<sup>1–6</sup>. Studies in which liquid flow is measured directly suggest that the viscosity of aqueous electrolytes confined to films of thickness greater than about 2–3 nm remains close to that in the bulk<sup>7–9</sup>; this behaviour is similar to that of non-associative organic liquids confined to films thicker than about 7–8 molecular layers<sup>8,10,11</sup>. Here we observe that the effective viscosity of water remains within a factor of three of its bulk value, even when it is confined to films in the thickness range  $3.5 \pm 1$  to  $0.0 \pm 0.4$  nm. This contrasts markedly with the behaviour of organic solvents, whose viscosity diverges when confined to films thinner than about 5–8 molecular layers<sup>10–15</sup>. We attribute this to the fundamentally different mechanisms of solidification in the two cases. For non-associative liquids, confinement promotes solidification by suppressing translational freedom of the molecules<sup>11,15–18</sup>; however, in the case of water, confinement seems primarily to suppress the formation of the highly directional hydrogen-bonded networks associated with freezing<sup>1,3</sup>.

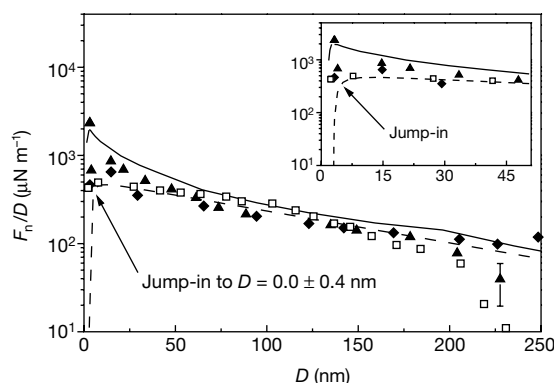
Highly purified water (conductivity water, see Fig. 1) was introduced between curved mica surfaces in a surface-force balance (SFB) capable of measuring both normal and lateral surface forces with extreme sensitivity and resolution<sup>11</sup>. Normal surface forces  $F_n(D)$ , where  $D$  is separation, were measured as a control both at the start and at several points during experiments, and are shown in Fig. 1. They reveal the long-ranged repulsion associated with the presence of charge on the mica surfaces, arising from the loss of potassium ions to solution. No indication of short-ranged hydration repulsion is seen, and, in agreement with earlier work<sup>19</sup>, the surfaces jump from separations  $D = D_j = 3.5 \pm 1$  nm into flat adhesive contact,  $D = D_0 = 0.0 \pm 0.4$  nm, as revealed by the position and shape of the optical interference fringes in the SFB. (As the size of a water molecule is roughly 0.25 nm, the presence of, at most, one monolayer of water per mica surface following adhesive contact cannot be ruled out within our resolution.) The spontaneous inward motion, which occurs in a single monotonic jump from  $D_j$  to  $D_0$ , is driven by van der Waals attraction between the surfaces overcoming the double-layer repulsion. It is due to an instability expected whenever  $|\partial F_n/\partial D| > K_n$ , where  $K_n$  is the constant of the normal spring (top inset to Fig. 2). The results reported here are based on seven separate experiments (using different pairs of mica sheets) as well as different contact points within experiments.

The surfaces were separated and enabled to approach slowly from

\* Present addresses: Service Recherche Technologies et Systèmes, LEGRAND SA 128, avenue de Lattre de Tassigny, 87045 Limoges Cedex, France (P.L.); and at Physical and Theoretical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, UK (J.K.).



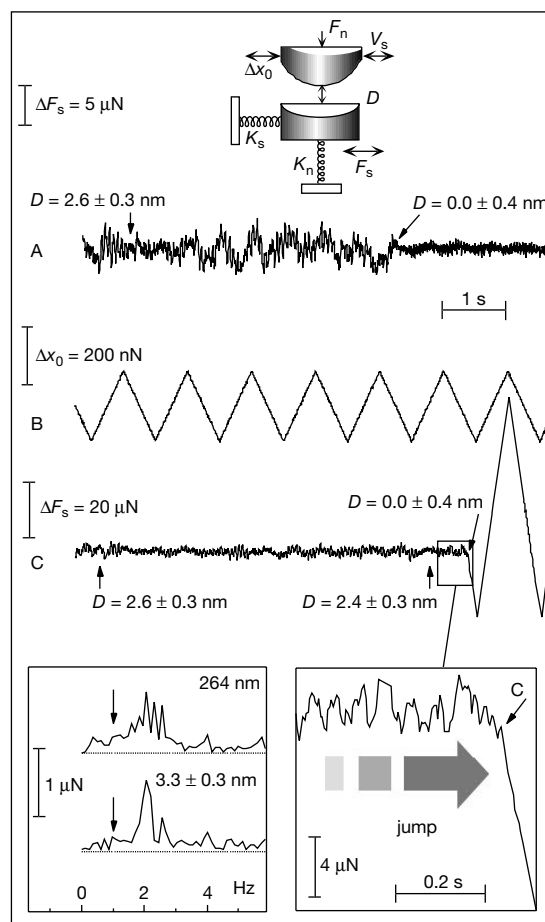
large separations,  $D > 10$  nm, under thermal drift (this varies slightly but was typically  $0.03 \text{ nm s}^{-1}$ ), while we recorded the lateral motion of the lower surface due to ambient vibrations, as shown in trace A of Fig. 2. Once the surfaces jump (from  $D = D_j$ ) into adhesive contact, the noise level drops markedly. The period  $\tau_j$  over which the jump occurs is estimated from observation of the interference fringes to be in the range  $0.2\text{--}0.5$  s, in line with earlier estimates<sup>20</sup>. In a different type of measurement, with the surfaces initially separated, the upper mica surface was made to move laterally back-and-forth parallel to the lower one at velocity  $v_s$  (trace B in Fig. 2). The shear forces  $F_s$  transmitted across the intersurface gap are simultaneously recorded as the surfaces approach under slow thermal drift. No shear forces greater than the noise-limited sensitivity  $\delta F_s$  ( $\pm 20$  nN, lower left inset) can be observed down to  $D = D_j$ , when the jump into adhesive contact results in rigid coupling between the surfaces (trace C, and lower right inset). In all cases the behaviour described was fully reproducible on subsequent separations and approaches. An upper limit  $\eta_{\text{eff}}^u$  on the effective shear viscosity may be estimated from the magnitude of the shear force  $F_s < \delta F_s$  transmitted across the gap<sup>11</sup> as  $\eta_{\text{eff}}^u = (\partial F_s / \partial D) / [(16\pi/5)R(v_s/D)]$ . This yields  $\eta_{\text{eff}}^u < 0.1 \text{ Pa s}$ , just before the jump in at  $D = D_j = 2.4 \pm 0.3$  nm for the run shown in Fig. 2. Although much larger than the viscosity of bulk water, this value is an upper limit and is consistent with earlier reports<sup>7–9</sup> on the fluidity of aqueous salt solutions confined to films thicker than about 2 nm. It is particularly noteworthy that during the jump itself, as clearly seen in the magnified region of trace C, the magnitude of  $F_s$  remains within the noise level right up to the point at which the



**Figure 1** Normal surface interaction measurements used to control for the water and surface characteristics in our experiments. Force ( $F_n/R$ ) against distance ( $D$ ) profile between curved mica surfaces (mean radius of curvature  $R = 1$  cm) across conductivity water at  $23 \pm 1$  °C. The solid curve is from the study on conductivity water by Pashley<sup>19</sup>, and the broken line corresponds to an exponential decay length (Debye length) of  $120 \pm 20$  nm, corresponding to  $(6.4 \pm 1) \times 10^{-6} \text{ M}$  1 : 1 salt concentrations and a surface potential of  $130 \pm 20$  mV. A solid–liquid surface tension  $-2.6 \pm 0.4 \text{ mN m}^{-1}$  can be deduced from the force required for pull-off from adhesive contact. These values are slightly lower than earlier reports<sup>19</sup>. The zero of the surface separation axis,  $D = 0$  in water, is at a position  $-0.5 \pm 0.3$  nm with respect to air contact between the surfaces, based on measurements at eight contact positions from seven different experiments, showing that a water-soluble layer coats the mica surfaces in air before addition of the conductivity water. Comparison with earlier studies<sup>20,28</sup> indicates that the adhesive contact separation in water in our experiments ( $D_0$ ) is, within our resolution, the contact separation of the mica lattice planes in the uncleaved crystals. Purification: tap water treated with activated charcoal was passed through a Millipore purification system (RiOs™ followed by a Milli-Q™ Gradient stage), yielding water with specific resistivity greater than  $18.2 \text{ M}\Omega$  (corresponding to less than  $4 \times 10^{-7} \text{ M}$  1 : 1 salt concentration) and total organic content less than 4 p.p.b. The higher ion concentration indicated by the measured Debye length is probably due to ions leached from the glassware and to dissolved  $\text{CO}_2$ . Different symbols correspond to different experiments. The inset shows the jump-in regime on an expanded scale.

surfaces come into rigid adhesive contact. This suggests that the viscosity of the water when confined—during the jump itself—to layers of thickness between  $3.5 \pm 1$  nm and  $0 \pm 0.4$  nm remains well within this upper limit  $\eta_{\text{eff}}^u$ .

More stringent limits on the viscosity of the water when it is confined to such subnanometre films ( $D < 2\text{--}3$  nm), down to the final few ångströms, may be set by considering in detail the processes occurring as the surfaces jump from  $D_j$  to adhesive contact. In this



**Figure 2** Measurements of the shear forces between sliding mica surfaces across water during approach to contact. The top cartoon shows schematically the relative motion, separation and direction of the forces measured between the two crossed cylindrical surfaces of mica in the surface force balance<sup>11</sup> (indicating the shear and normal force springs of respective constants  $K_s$  and  $K_n$ ). The top surface may be moved laterally in the  $\pm \Delta x_0$  direction by means of a sector piezoelectric tube (PZT), while the distortion  $\Delta x$  of the shear spring gives  $F_s = K_s \Delta x$  and is monitored directly from changes in an air-gap capacitor. Trace A shows the noise in  $F_s$  arising from ambient vibrations as the surfaces approach each other slowly (at about  $0.03 \text{ nm s}^{-1}$ ) under thermal drift and jump into contact to  $D = D_0 = 0.0 \pm 0.4$  nm (arrow). Trace B shows the back-and-forth lateral motion applied to the top mica surface by the sector PZT as the surfaces drift slowly towards each other. Trace C shows the corresponding shear force  $F_s$  transmitted between the surfaces: it is indistinguishable from the noise until they have jumped together into adhesive contact (arrow and point C on the expanded lower right inset), following which  $F_s = K_s \Delta x = K_s \Delta x_0$  as the two surfaces move in tandem. Frequency analysis (lower left inset) reveals that the magnitude of  $F_s$  at the frequency of the lateral drive motion (1 Hz, indicated by arrow) for a surface separation  $D = 3.3$  nm is essentially identical to its value at large separations ( $D = 264$  nm), both being due to bending of the connecting PZT wires<sup>11</sup> and to ambient noise rather than to any transmitted shear force  $F_s$ . Their difference ( $0 \pm 20$  nN, averaged over several experiments) may be taken as the noise-limited sensitivity  $\delta F_s$  in measuring the shear force between the surfaces.

regime, if we treat the effective geometry of the curved surfaces as that of a sphere of radius  $R$  approaching a flat surface, the equation of motion is to a good approximation

$$M(d^2D/dt^2) + 6\pi R^2\eta_{\text{eff}}[(dD/dt)/D] = F_{\text{vdW}}(D) - K_n(D_j - D) \quad (1)$$

where  $M$  is the effective mass of the moving mica surface and its mounting, the second term on the left is the Reynolds hydrodynamic resistance<sup>21</sup> to the approach of a sphere to a flat across a medium of viscosity  $\eta_{\text{eff}}$ , and we ignore the very small change in the double-layer interaction in the interval  $\{D_j, D_0\}$ . In this surface separation regime the dominant force driving the approach of the surfaces is the van der Waals dispersive attraction,  $F_{\text{vdW}}(D) = AR/6D^2$ , where  $A = -2 \times 10^{-20}$  J is the relevant Hamaker constant. It is readily shown that in the conditions of our experiments the inertial term and that (in  $K_n$ ) due to the spring bending are negligible compared with the other terms. The time  $\tau_j$  for the surfaces to jump from  $D_j$  to  $D_0$  is thus

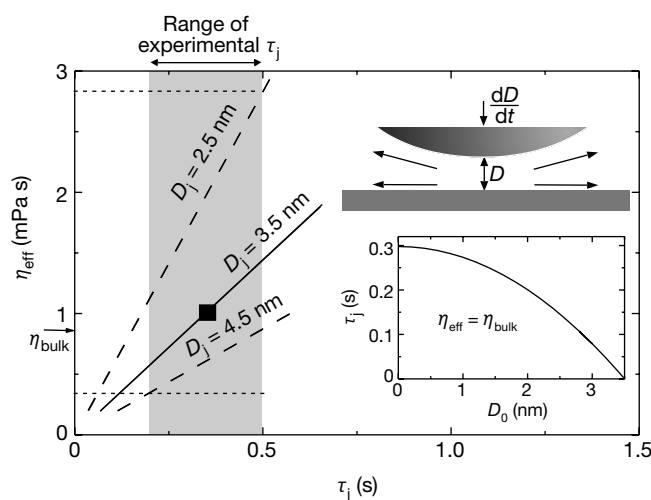
$$\tau_j(\eta_{\text{eff}}, D_j, D_0) = \int_{D_0}^{D_j} \frac{36\pi R\eta_{\text{eff}}D}{A} dD = \left(\frac{18\pi R\eta_{\text{eff}}}{A}\right)(D_j^2 - D_0^2) \quad (2)$$

The variation of  $\tau_j$  with  $D_0$  is shown in the inset to Fig. 3 for  $D_j = 3.5$  nm and  $\eta_{\text{eff}} = \eta_{\text{bulk}} = 0.86 \times 10^{-3}$  Pa s, the viscosity of bulk water at 23 °C. The vertical shaded band in Fig. 3 shows the range of experimental  $\tau_j$  values; also shown, for values of the jump separation in the range  $D_j = (2.5\text{--}4.5)$  nm, is the implicit relation (from equation (2)) between the jump time to  $D_0 = 0$  and the effective viscosity  $\eta_{\text{eff}}$ . The range of  $\eta_{\text{eff}}$  where  $\tau_j(\eta_{\text{eff}}, D_j, 0)$  overlaps the shaded band, shown by horizontal dotted lines, thus defines the range of effective viscosities that correspond to the experimental jump times. This shows that the viscosity of water confined to films of thickness in the range  $3.5 \pm 1$  nm down to  $0.0 \pm 0.4$  nm is within a factor of 3 or so in either direction of the viscosity of bulk water  $\eta_{\text{bulk}}$ . The uncertainty arises from the scatter in the values of the jump distance  $D_j$  and the jump time  $\tau_j$ . It is of interest that the viscosity corresponding to the mid-range of both  $D_j$  (3.5 nm) and the experimental  $\tau_j$  (0.35 s), indicated as the filled square in Fig. 3, is close to that of bulk water. A more detailed consideration shows that the magnitude of  $\eta_{\text{eff}}$  remains within or close to this range at all film

thicknesses throughout the jump into contact. We note also that the effective maximal shear rates during our measurements vary from about  $10^2 \text{ s}^{-1}$  in the absence of applied lateral motion before the jump-in (Fig. 2a), to about  $10^4 \text{ s}^{-1}$  during the jump-in itself<sup>22</sup>. Within our resolution, the effective viscosity of the confined water is similar at these different shear rates, indicating that its behaviour is Newtonian within this range.

This result represents a direct evaluation of the viscosity of water confined to subnanometre films, and is in remarkable contrast to the properties of similarly confined non-associative organic liquids. Experiments and simulations show that both water and organic liquids undergo layering near smooth solid surfaces, and that both retain their bulk fluidity when confined to films thicker than some 8–10 molecular layers<sup>6–11,14,15,23</sup>. Simple non-associative liquids, however, are solid-like or have diverging viscosities when they are confined to films less than 5–8 molecular layers thick<sup>10–15</sup>, whereas the viscosity of water, as revealed here, remains comparable to its bulk value even within films down to one or two monolayers thick<sup>24</sup>.

Why does water behave so differently from organic liquids confined to correspondingly thin films between solid surfaces? Models for non-associative liquids (which include most organic solvents) suggest that, as the confinement increases, the translational entropy available to the molecules decreases to the point at which it becomes thermodynamically favourable for them to condense to an ordered phase<sup>15–18</sup>, with consequent solid-like behaviour even at temperatures where the bulk is disordered and fluid. The origin of solidification in water is rather different: the formation of ice with its highly directional hydrogen-bond lattice<sup>1,3</sup> is thermodynamically favoured despite the large accompanying loss of the hydrogen-bond entropy available in liquid water. Now we expect that the confining surfaces locally must induce translational order (because they reduce the translational entropy of adjacent water molecules much as for non-associative molecules). Moreover, the mica surface is hydrophilic and fairly commensurate with water<sup>25</sup> and this should provide an additional surface ordering mechanism. Thus our observation that the confined water nonetheless retains its bulk fluidity presumably implies that these ordering effects cannot



**Figure 3** Evaluation of the effective viscosity of confined water from the jump-in data (Fig. 2). The cartoon (top right) illustrates the jump-in driven by van der Waals attraction and opposed by the hydrodynamic force resulting from the expulsion of the viscous liquid from between the surfaces. The inset (lower right) shows how, for an effective viscosity  $\eta_{\text{eff}} = \eta_{\text{bulk}} = 0.86$  mPa s (the viscosity of bulk water at 23 °C), the jump time  $\tau_j$  varies with jump distance for a jump commencing at  $D_j = 3.5$  nm, according to equation (2). The main figure shows the effective viscosity  $\eta_{\text{eff}}$  of the liquid being squeezed out of the gap corresponding to the jump time to contact  $\tau_j$ , for three values of the

jump distance  $D_j$ : the mid-range value 3.5 nm (solid line) and the two extremes from which jumps occurred,  $D_j = 2.5$  and 4.5 nm (broken lines). The shaded vertical band is the range of experimentally determined  $\tau_j$ , so that the range of effective viscosity  $\eta_{\text{eff}}$  of the confined water films in the thickness interval  $(3.5 \pm 1 \text{ nm}, 0)$ , which corresponds to the experiments, is defined (horizontal dotted lines) by the overlap of the broken lines with the shaded band. This possible range of values is within a factor of 3 or so either side of  $\eta_{\text{bulk}}$ ; its mid-range value with respect to  $D_j$  and  $\tau_j$ , indicated as a solid square, is close to  $\eta_{\text{bulk}}$  itself.

overcome the network entropy of the hydrogen bonding in the fluid state. Equivalent differences are found also in the bulk: above a critical density or pressure, non-associative liquids spontaneously go to an ordered phase, whereas it is well known that pressure promotes melting of ice. The effect of confinement is in some measure equivalent to a pressure (or density) increase relative to the unconfined bulk, due to the attraction exerted on the solvent molecules by the confining walls<sup>26</sup>; such a density increase opposes any tendency to freeze. This suggests that the fluidity of water even under the most severe confinement is directly related to the well-known anomalous density decrease of water on freezing, as well as to the considerable supercooling of water possible near surfaces and in small droplets<sup>1</sup>.

This persistent fluidity clearly has interesting consequences for many phenomena where confined water is implicated. It should be emphasized, however, that our present study examined conductivity water, where the ion concentration (of the order of  $10^{-6}$  M) is lower by many orders of magnitude than is commonly found in nature or technology and where the confining surfaces attract each other at subnanometre separations. It is possible that where the surfaces are covered with hydrated ions (as for the case of mica across electrolytes at higher salt concentrations), and thus repel at such separations, the behaviour of the confined aqueous films may be different<sup>24,27</sup>. □

Received 6 April; accepted 24 June 2001.

1. Clifford, J. in *Water in Disperse Systems* (ed. Franks, F.) 75–132 (Plenum, New York and London, 1975).
2. Drake, J. M. & Klafter, J. Dynamics of confined molecular systems. *Phys. Today* **43**, 43–45 (1990).
3. Bellissent-Funel, M.-C. & Dore, J. C. (eds) *Hydrogen Bond Networks* (NATO ASI Series) (Kluwer Academic, Dordrecht, 1994).
4. Xia, X., Perera, L., Essmann, U. & Berkowitz, M. L. The structure of water at platinum/water interfaces: Molecular dynamics computer simulations. *Surf. Sci.* **335**, 401–415 (1995).
5. Meyer, M. & Stanley, H. E. Liquid–liquid phase transition in confined water: A Monte Carlo study. *J. Phys. Chem. B* **103**, 9728–9730 (1999).
6. Gallo, P., Rovere, M. & Spohr, E. Glass transition and layering effects in confined water: A computer simulation study. *J. Chem. Phys.* **113**, 11324–11335 (2000).
7. Roberts, A. D. & Tabor, D. The extrusion of liquids between highly elastic solids. *Proc. R. Soc. Lond. A* **325**, 323–345 (1971).
8. Israelachvili, J. N. Measurement of the viscosity of liquids in very thin films. *J. Colloid Interf. Sci.* **110**, 263–271 (1986).
9. Horn, R. G., Smith, D. T. & Haller, W. Surface forces and viscosity of water measured between silica sheets. *Chem. Phys. Lett.* **162**, 404–408 (1989).
10. Granick, S. Motions and relaxations of confined liquids. *Science* **253**, 1374–1379 (1991).
11. Klein, J. & Kumacheva, E. Simple liquids confined to molecularly thin layers. I. Confinement-induced liquid to solid phase transitions. *J. Chem. Phys.* **108**, 6996–7009 (1998).
12. Rhykerd, C. L., Schoen, M., Diestler, D. J. & Cushman, J. H. Epitaxy in simple classical fluids in micropores and near-solid surfaces. *Nature* **330**, 461–463 (1987).
13. Israelachvili, J., McGuiggan, P. M. & Homola, A. M. Dynamic properties of molecularly thin liquid films. *Science* **240**, 189–191 (1988).
14. Thompson, P. A., Robbins, M. O. & Grest, G. S. Structure and shear response in nanometer-thick films. *Isr. J. Chem.* **35**, 93–106 (1995).
15. Gao, J., Luedtke, W. D. & Landman, U. Layering transitions and dynamics of confined liquid films. *Phys. Rev. Lett.* **79**, 705–708 (1997).
16. Chandler, D., Weeks, J. D. & Anderson, H. C. Van der Waals picture of liquids, solids and phase transformations. *Science* **220**, 787–794 (1983).
17. Tkachenko, A. & Rabin, Y. Effect of boundary conditions in fluctuations and solid-liquid transition in confined films. *Langmuir* **13**, 7146–7150 (1997).
18. Weinstein, A. & Safran, S. A. Surface and bulk ordering in thin films. *Europhys. J.* **42**, 61–64 (1998).
19. Pashley, R. M. Hydration forces between mica surfaces in aqueous electrolyte solutions. *J. Colloid Interf. Sci.* **80**, 153–162 (1980).
20. Israelachvili, J. N. & Adams, G. E. Measurement of forces between two mica surfaces in aqueous electrolyte solutions in the range 0–100 nm. *J. Chem. Soc. Faraday Trans. 1* **79**, 975–1001 (1978).
21. Happel & Brenner, H. *Low Reynolds Number Hydrodynamics* (Prentice Hall, Englewood Cliffs, 1965).
22. Chan, D. Y. C. & Horn, R. G. The drainage of thin liquid films between solid surfaces. *J. Chem. Phys.* **83**, 5311–5324 (1985).
23. Pashley, R. M. & Israelachvili, J. N. Molecular layering of water in thin films between mica surfaces and its relation to hydration forces. *J. Colloid Interf. Sci.* **101**, 511–523 (1984).
24. Homola, A. M., Israelachvili, J. N., Gee, M. L. & McGuiggan, P. M. Measurements of and relation between the adhesion and friction of two surfaces separated by molecularly thin liquid films. *J. Tribol.* **111**, 675–682 (1989).
25. Ravina, I. & Low, P. F. Relation between swelling, water properties and b-dimension in montmorillonite-water systems. *Clays Clay Minerals* **20**, 109–123 (1972).
26. Cui, S. T., Cummings, P. T. & Cochran, H. D. Molecular simulation of the transition from liquidlike to solidlike behavior in complex fluids confined to nanoscale gaps. *J. Chem. Phys.* **114**, 7189–7195 (2001).

27. Berman, A., Drummond, C. & Israelachvili, J. Amontons' Law at the molecular level. *Tribol. Lett.* **4**, 95–101 (1998).
28. Tabor, D. & Winterton, R. H. The direct measurement of normal and retarded van der Waals forces. *Proc. R. Soc. A* **312**, 435–450 (1969).

## Acknowledgements

We thank D. Chandler, J. Israelachvili, S. Safran and S. Titmuss for comments and discussions. We thank the Eshkol Foundation for a studentship (U.R.), and the US–Israel Binational science Foundation, the Deutsche–Israelische Program (DIP) and the Minerva Foundation for their support of this work.

Correspondence and requests for materials should be addressed to J.K. (e-mail: jacob.klein@weizmann.ac.il or jacob.klein@chem.ox.ac.uk).

## Anomalous properties in ferroelectrics induced by atomic ordering

A. M. George, Jorge Iñiguez & L. Bellaiche

Physics Department, University of Arkansas, Fayetteville, Arkansas 72701, USA

Complex insulating perovskite alloys are of considerable technological interest because of their large dielectric and piezoelectric responses. Examples of such alloys include  $(\text{Ba}_{1-x}\text{Sr}_x)\text{TiO}_3$ , which has emerged as a leading candidate dielectric material for the memory-cell capacitors in dynamic random access memories<sup>1</sup>; and  $\text{Pb}(\text{Zr}_{1-x}\text{Ti}_x)\text{O}_3$  (PZT), which is widely used in transducers and actuators<sup>2</sup>. The rich variety of structural phases that these alloys can exhibit, and the challenge of relating their anomalous properties to the microscopic structure, make them attractive from a fundamental point of view. Theoretical investigations of modifications to the atomic ordering of these alloys suggest the existence of further unexpected structural properties<sup>3</sup> and hold promise for the development of new functional materials with improved electromechanical properties. Here we report *ab initio* calculations that show that a certain class of atomic rearrangement should lead simultaneously to large electromechanical responses and to unusual structural phases in a given class of perovskite alloys. Our simulations also reveal the microscopic mechanism responsible for these anomalies.

The class of  $\text{Pb}(\text{Mg}_{1/3}\text{Nb}_{2/3})\text{O}_3$ – $\text{PbTiO}_3$  (PMN–PT) and  $\text{Pb}(\text{Zn}_{1/3}\text{Nb}_{2/3})\text{O}_3$ – $\text{PbTiO}_3$  (PZN–PT) perovskite ferroelectric alloys have been reported<sup>4</sup> to show remarkably large piezoelectric constants around  $2,000 \text{ pC N}^{-1}$ . These materials thus promise improvements in the resolution and range of ultrasonic and sonar listening devices<sup>5</sup>. Perovskite  $\text{A}(\text{B}'\text{B}'')\text{O}_3$  alloys are also of great interest, as demonstrated by the discovery of an unexpected monoclinic phase<sup>6</sup> in  $\text{Pb}(\text{Zr}_{1-x}\text{Ti}_x)\text{O}_3$ . These two findings have led to a search for materials with even larger electromechanical response and/or currently unobserved phases. In particular, if a mechanism can be found that occurs in a large number of ferroelectric alloys, markedly enhances the piezoelectricity and the dielectric responses, and leads to unexpected structural features, that mechanism is likely to have large technological and fundamental implications. Here we report that such a mechanism exists and simply consists in rearranging in a certain way the atoms in a heterovalent alloy (that is, an alloy made of atoms belonging to different columns of the periodic table). More precisely, we predict that materials made of the sequences  $\text{Pb}(\text{Sc}_{0.5+\nu}^{3+}\text{Nb}_{0.5-\nu}^{5+})\text{O}_3$ /  $\text{Pb}(\text{Sc}_{0.5}^{3+}\text{Nb}_{0.5}^{5+})\text{O}_3$ /  $\text{Pb}(\text{Sc}_{0.5-\nu}^{3+}\text{Nb}_{0.5+\nu}^{5+})\text{O}_3$ /  $\text{Pb}(\text{Sc}_{0.5}^{3+}\text{Nb}_{0.5}^{5+})\text{O}_3$  along the [001] direction should show very large electromechanical responses,



and unusual monoclinic and orthorhombic phases for some values of the  $\nu$  parameter. The analysis of our results points to the different valence of the B atoms (Sc and Nb) as the main reason for the existence of these anomalous properties.

Here, we use the numerical scheme proposed in refs 7 and 8, which consists of constructing an effective hamiltonian for the alloy from first-principles calculations, to predict the properties of  $\text{Pb}(\text{Sc}_{1-x}\text{Nb}_x)\text{O}_3$  (PSN) structures. This effective hamiltonian<sup>7,8</sup> contains a local-mode self-energy (expanded in even-order terms up to fourth order of the phonon local soft modes), a long-range dipole–dipole interaction, a short-range interaction between soft modes (quadratic order of the soft modes), an elastic energy (expanded to second order of the strain variables), and an interaction between the local modes and local strain (second order of the soft modes and first order in strain). It also contains an energy term describing the effects of the atomic configuration on the phonon local soft modes (which are directly related to the electrical polarization) which can be written as

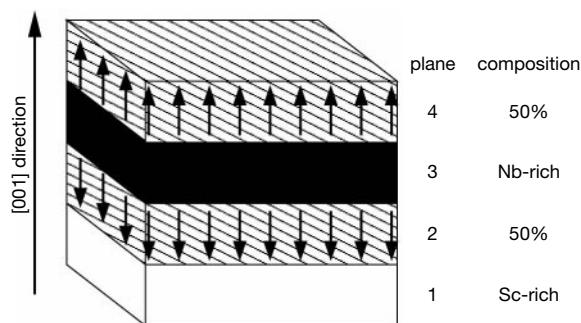
$$\Delta E = - \sum_i Z^* \mathbf{u}_i \cdot \left[ \sum_j - \sigma_j S_{ji} \hat{\mathbf{e}}_{ji} \right] \quad (1)$$

where  $i$  runs over all the cells and  $j$  runs over the three nearest-neighbour shells of cell  $i$ . Here  $\mathbf{u}_i$  is the local soft mode in cell  $i$ ,  $Z^*$  is the Born effective charge associated with the local soft modes, and  $\hat{\mathbf{e}}_{ji}$  is the unit vector joining the B site  $j$  to the B site  $i$ . The variables  $\{\sigma_j\}$  characterize the atomic configuration:  $\sigma_j = +1$  ( $-1$ ) indicates that there is an Nb (Sc) ion in cell  $j$ .  $S_{ji}$  is an alloy-related parameter that only depends on the distance between the B sites  $i$  and  $j$ , and is derived from first-principles calculations<sup>9–12</sup> on small cells. Equation (1) indicates that  $\Delta E$  can be viewed as the interaction energy between the dipole moment  $Z^* \mathbf{u}_i$  associated with the site  $i$  and an internal electric field  $\sum_j - \sigma_j S_{ji} \hat{\mathbf{e}}_{ji}$  induced by the B ions of sites  $j$  on the site  $i$ . Moreover, we numerically find that the  $S_{ji}$  parameters all have negative signs. As a result, the radial electric field  $-\sigma_j S_{ji} \hat{\mathbf{e}}_{ji}$  acting on site  $i$  and induced by a  $\text{Sc}^{3+}$  ( $\text{Nb}^{5+}$ ) ion sitting on site  $j$  is directed from the site  $i$  to the site  $j$  (respectively, from site  $j$  to site  $i$ ). This is consistent with an electrostatic picture of PSN since Sc (Nb) ions are negatively (positively) charged with respect to the average B-ion valence of  $4+$ . The effective hamiltonian approach yields good agreement with direct first-principles results in PZT and PSN alloys<sup>7,8</sup>. Previous works (see refs 7–8) have found that a linear rescaling of the simulation temperature often leads to good agreement with experiment. Adopting this approach again here, our temperatures are rescaled down by a factor of 2.5 so that the theoretical Curie temperature in disordered PSN is forced to match the experimental value<sup>13</sup>. The need for such rescaling may be due to higher perturbative terms or the rotation of the oxygen

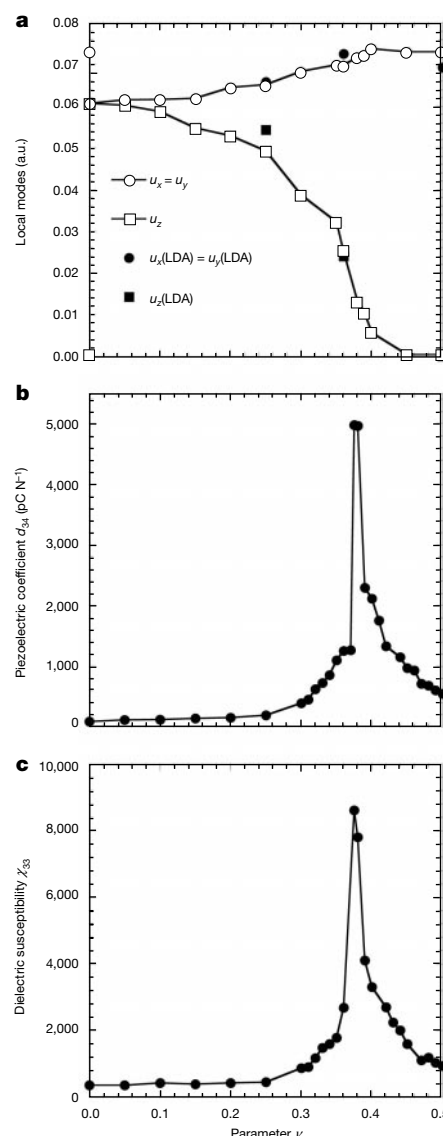
octahedra, neglected in our effective model of PSN.

Here, we use the total energy of our alloy effective hamiltonian in Monte Carlo simulations to calculate finite-temperature properties of some selected  $\text{Pb}(\text{Sc}_{1-x}\text{Nb}_x)\text{O}_3$  structures. We consider structures with the following sequence of four B-planes along the  $[001]$  direction (Fig. 1): a niobium-poor plane for which  $x = 0.5 - \nu$ , a second plane made of 50% scandium and 50% niobium ( $x = 0.5$ ), a niobium-rich plane with  $x = 0.5 + \nu$ , and a fourth plane similar to the second one. The B atoms are randomly distributed within each of these four planes. The studied structures only differ in the value of the parameter  $\nu$ , which is allowed to vary from zero (in the case of the disordered PSN alloy) to 0.5 (in the case in which the first plane is entirely made of Sc atoms and the third plane is fully occupied by Nb atoms). We use  $12 \times 12 \times 12$  supercells, implying that the sequence of the four different B(001) planes is repeated three times, to get well converged results<sup>7,8</sup>.

Figure 2a shows the cartesian coordinates ( $u_x$ ,  $u_y$  and  $u_z$ )—along

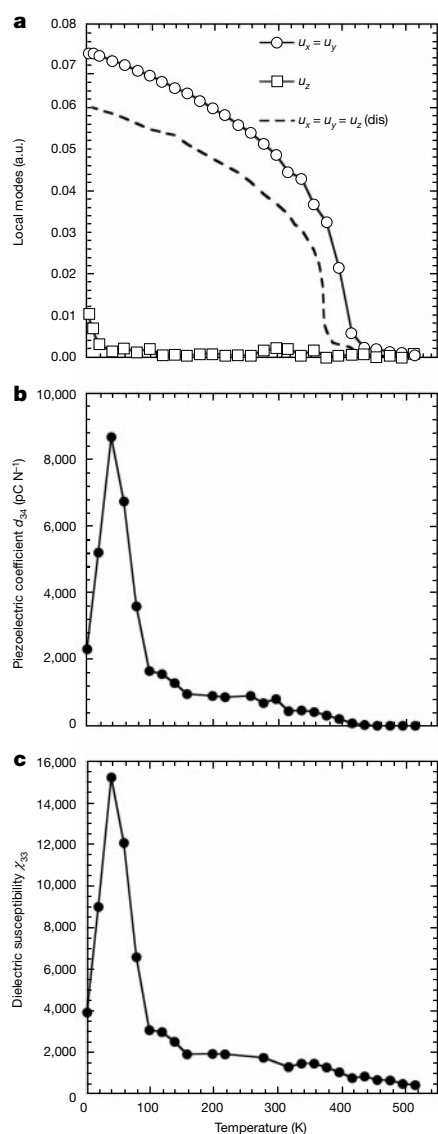


**Figure 1** Schematic representation of the ferroelectric structures considered here. The materials are  $\text{Pb}(\text{Sc}_{0.5+\nu}^{3+}\text{Nb}_{0.5-\nu}^{5+})\text{O}_3/\text{Pb}(\text{Sc}_{0.5}^{3+}\text{Nb}_{0.5}^{5+})\text{O}_3/\text{Pb}(\text{Sc}_{0.5-\nu}^{3+}\text{Nb}_{0.5+\nu}^{5+})\text{O}_3/\text{Pb}(\text{Sc}_{0.5}^{3+}\text{Nb}_{0.5}^{5+})\text{O}_3$  structures. The internal electric fields acting on the four B planes are indicated by arrows.



**Figure 2** Properties of the studied structures as a function of the  $\nu$  parameter at 20 K. **a**, Average cartesian coordinates  $u_x$ ,  $u_y$  and  $u_z$  of the local mode; **b**, the  $d_{34}$  piezoelectric coefficient; **c**, the  $\chi_{33}$  dielectric susceptibility. The filled symbols in **a** represent the cartesian coordinates  $u_x(\text{LDA})$ ,  $u_y(\text{LDA})$  and  $u_z(\text{LDA})$  of the local modes predicted by direct first-principles calculations made using the local density approximation<sup>9–12</sup>. a.u., arbitrary units.

the [100], [010] and [001] directions, respectively—of the supercell average of the local mode vectors as a function of this parameter  $\nu$  at 20 K. The structure for which  $\nu$  is null has  $u_x = u_y = u_z \neq 0$ . This characterizes a ferroelectric rhombohedral structure in which the polarization is directed along the pseudo-cubic [111] direction, consistent with experiments on disordered PSN samples<sup>13</sup>. Interestingly, increasing  $\nu$  results in a strong decrease of  $u_z$  whereas  $u_x$  and  $u_y$  slowly increase and remain equal to each other. This behaviour corresponds to a ferroelectric phase of monoclinic symmetry for which the polarization lies between the [111] and [110] directions (the one denoted  $M_B$  in ref. 14). When  $\nu$  is larger than 0.44,  $u_z$  becomes null and  $u_x$  and  $u_y$  reach their maximum value, indicating that the resulting phase is now of orthorhombic symmetry with a polarization lying along the [110] direction. Increasing the parameter  $\nu$  thus leads to three different ground states and to a continuous rotation of the electrical polarization. To our knowledge, the predicted monoclinic  $M_B$  ground state has never been observed in any perovskite material without the aid of external stress



**Figure 3** Properties of the structure associated with  $\nu = 0.375$  as a function of temperature. Panels **a**, **b** and **c** display the same property as the corresponding panels of Fig. 2. Dashed line in **a** shows one of the three cartesian coordinates ( $u_k$ ) of the local mode in disordered (dis) PSN (the two other cartesian coordinates are nearly identical to the one displayed, and are omitted for clarity).

or electric field. Similarly, we are not aware of any studies pointing to an orthorhombic ground state in PSN. The existence of these monoclinic and orthorhombic phases is due to the energy term described in equation (1). Figure 2a also shows the predictions of local density approximation<sup>9–11</sup> calculations performed on 20-atom supercells mimicking the structures corresponding to  $\nu = 0.25$ ,  $\nu = 0.36$  and  $\nu = 0.50$ . In these supercells, each B(001) plane is represented by a virtual atom corresponding to the composition in this plane<sup>12,15</sup>. The direct first-principles calculations agree with the predictions of our effective hamiltonian approach, thus further supporting the existence of these unusual monoclinic and orthorhombic phases.

Figure 2b summarizes the effects of the parameter  $\nu$  on piezoelectricity and Fig. 2c the effects on dielectric response at 20 K. We found that the electromechanical coefficients most affected by the atomic ordering are the shear  $d_{34}$  piezoelectric coefficient and the  $\chi_{33}$  dielectric susceptibility, when expressing both piezoelectric and dielectric tensors in the orthonormal basis formed by  $\mathbf{a}_1 = [100]$ ,  $\mathbf{a}_2 = [010]$  and  $\mathbf{a}_3 = [001]$ . More precisely, Fig. 2b shows that  $d_{34}$  reaches a peak and remains at large values for a broad range of  $\nu$  centred around the monoclinic-to-orthorhombic phase transition. Figure 2c demonstrates that the studied atomic ordering simultaneously results in a large dielectric response, as  $\chi_{33}$  achieves values above 1,000 at 20 K when  $\nu$  is greater than 0.3.

We now investigate the finite-temperature properties of a structure showing one of the largest electromechanical responses at 20 K. More precisely, we focus on the structure for which  $\nu = 0.375$ . Figure 3a shows the cartesian coordinates of the supercell average of the local mode vectors as a function of temperature for this structure, and compares them with those of the disordered PSN material. Each coordinate of the local mode in each structure is close to zero at high temperature, characterizing a paraelectric phase. At a temperature close to 373 K, the disordered material undergoes a transition from a paraelectric cubic phase to a ferroelectric rhombohedral structure (for which  $u_x = u_y = u_z$ ), consistent with experiments<sup>13</sup>. The disordered PSN material then remains in the rhombohedral phase for lower temperature. On the other hand, the modulated structure with  $\nu = 0.375$  adopts three different phases: a paraelectric tetragonal phase induced by atomic ordering (for which  $u_x = u_y = u_z = 0$ ) at high temperature, an orthorhombic ferroelectric phase ( $u_x = u_y \neq 0$  and  $u_z = 0$ ) for temperature between 400 K and 40 K, and the monoclinic ferroelectric  $M_B$  phase ( $u_x = u_y > u_z$ , with  $u_z \neq 0$ ) for temperature lower than 40 K. As shown in Fig. 3b and c, the existence of the orthorhombic-to-monoclinic phase transition results in huge electromechanical responses peaking around this transition and occurring over a broad range of temperature. In fact,  $d_{34}$  and  $\chi_{33}$  are greater than 1,500 pC N<sup>-1</sup> and 3,000, respectively, for any temperature lower than 100 K.

We also numerically found (results not shown here) that continuously increasing  $\nu$  from 0 to 0.44 leads to a continuous decrease of the orthorhombic-to-monoclinic transition temperature from 373 K to 0 K. As a result, the temperature at which  $d_{34}$  and  $\chi_{33}$  both peak depends on the value of the  $\nu$  parameter. This dependency could lead to the development of devices with electromechanical performances optimized for any temperature between 373 K and 0 K.

The intriguing results of Figs 2 and 3 can be understood by means of equation (1). The internal electric field  $\Sigma_j - \sigma_j S_{ji} \hat{\mathbf{e}}_{ji}$  acting on each B(001) plane is represented in Fig. 1. The atoms in plane 2 feel an internal electric field oriented along the  $[00\bar{1}]$  direction whereas the atoms in plane 4 feel an internal electric field along the  $[001]$  direction. It is straightforward to demonstrate that the magnitude of these fields is linearly dependent on the parameter  $\nu$ . (These features can be understood qualitatively by simple short-range electrostatic considerations: the difference of valence between Sc and Nb atoms leads to charged B(001) planes with ionic charges

linearly dependent on  $\nu$ . When electrostatic interactions up to the third-neighbour shells are included, these charged planes generate internal electric fields whose directions are those of Fig. 1 and whose magnitudes are linearly dependent on  $\nu$ .) Increasing  $\nu$  thus results in stronger internal fields with opposite directions. These strong opposite fields tend to suppress the polarization's component along the direction of compositional modulation. As a result,  $u_z$  becomes smaller than  $u_x$  and  $u_y$ . The studied structures with intermediate values of  $\nu$  thus first undergo a paraelectric-to-orthorhombic ferroelectric transition at high temperature before adopting the monoclinic  $M_b$  phase for ground state. The modulated structures with the largest values of  $\nu$  have the strongest internal electric fields, which annihilate  $u_z$  at any temperature. Consequently, such structures never reach the monoclinic phase and instead adopt the orthorhombic phase for ground state. The large values of  $d_{34}$  and  $\chi_{33}$  shown in Figs 2 and 3 simply reflect the considerable change of  $u_z$  when some parameters are slightly modified ( $\nu$  at a fixed temperature, or  $T$  for a fixed value of  $\nu$ ), especially for structures at the borderline between the monoclinic and orthorhombic phases. In other words, the large electromechanical responses are consistent with the ease of rotating the polarization<sup>7,16</sup>. We have shown here results for the thinnest possible structures. For larger structures, with thicker layers, the 'unusual' electromechanical responses will be smaller due to smaller internal electric fields.

On the basis of the electrostatic considerations discussed above, we expect that any alloy made of heterovalent atoms and with a rhombohedral ground state in its disordered form should have the structural, piezoelectric and dielectric anomalies displayed in Figs 2 and 3, when the atomic ordering along the [001] direction is adjusted in a certain way. The atomically ordered structures discussed here could be grown by means of a pulse laser deposition technique<sup>17</sup> or by using molecular beam epitaxy. □

Received 4 April; accepted 29 June 2001.

1. Abe, K. & Komatsu, S. Ferroelectric properties in epitaxially grown  $\text{Ba}_x\text{Sr}_{1-x}\text{TiO}_3$  thin films. *J. Appl. Phys.* **77**, 6461–6465 (1995).
2. Uchino, K. *Piezoelectric Actuators and Ultrasonic Motors* (Kluwer Academic, Boston, 1996).
3. Sai, N., Meyer, B. & Vanderbilt, D. Compositional inversion symmetry breaking in ferroelectric perovskites. *Phys. Rev. Lett.* **84**, 5636–5639 (2000).
4. Park, S.-E. & Shrout, T. R. Ultrahigh strain and piezoelectric behavior in relaxor based ferroelectric single crystal. *J. Appl. Phys.* **82**, 1804–1811 (1997).
5. Service, R. F. Shape-changing crystals get shifter. *Science* **275**, 1878 (1997).
6. Noheda, B. *et al.* A monoclinic ferroelectric phase in the  $\text{Pb}(\text{Zr}_{1-x}\text{Ti}_x)\text{O}_3$  solid solution. *Appl. Phys. Lett.* **74**, 2059–2061 (1999).
7. Bellaiche, L., Garcia, A. & Vanderbilt, D. Finite-temperature properties of  $\text{Pb}(\text{Zr}_{1-x}\text{Ti}_x)\text{O}_3$  alloys from first principles. *Phys. Rev. Lett.* **84**, 5427–5430 (2000).
8. Hemphill, R., Bellaiche, L., Garcia, A. & Vanderbilt, D. Finite-temperature properties of disordered and ordered  $\text{Pb}(\text{Sc}_{0.5}\text{Nb}_{0.5})\text{O}_3$  alloys. *Appl. Phys. Lett.* **77**, 3642–3644 (2000).
9. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–871 (1964).
10. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–1138 (1965).
11. Vanderbilt, D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B* **41**, 7892–7895 (1990).
12. Bellaiche, L. & Vanderbilt, D. Virtual crystal approximation revisited: Application to dielectric and piezoelectric properties of perovskites. *Phys. Rev. B* **61**, 7877–7882 (2000).
13. Chu, F., Reaney, I. M. & Setter, N. Spontaneous (zero-field) relaxor-to-ferroelectric-phase transition in disordered  $\text{Pb}(\text{Sc}_{1/2}\text{Nb}_{1/2})\text{O}_3$ . *J. Appl. Phys.* **77**, 1671–1676 (1995).
14. Vanderbilt, D. & Cohen, M. H. Monoclinic and triclinic phases in higher-order Devonshire theory. *Phys. Rev. B* **63**, 094108 (2001).
15. Ramer, N. J. & Rappe, A. M. Application of a new virtual crystal approach for the study of disordered perovskites. *J. Phys. Chem. Solids* **61**, 315–320 (2000).
16. Fu, H. & Cohen, R. E. Polarization rotation mechanism for ultrahigh electromechanical response in single crystal piezoelectrics. *Nature* **403**, 281–283 (2000).
17. Brazier, M., McElfresh, M. & Mansour, S. Unconventional hysteresis behavior in compositionally graded  $\text{Pb}(\text{Zr,Ti})\text{O}_3$  thin films. *Appl. Phys. Lett.* **72**, 1121–1123 (1998).

## Acknowledgements

This work was supported by the National Science Foundation, the Office of Naval Research and the Arkansas Science and Technology Authority. We thank D. Vanderbilt and A. Garcia for discussions.

Correspondence and requests for materials should be addressed to L.B.  
(e-mail: laurent@comp.uark.edu).

# Elasticity of iron at the temperature of the Earth's inner core

Gerd Steinle-Neumann\*, Lars Stixrude\*, R. E. Cohen† & Oguz Gülseren‡

\* Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA

† Carnegie Institution of Washington and Center for High Pressure Research, Washington DC 20015-1305, USA, and Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

‡ NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899-8562 USA, and Department of Material Sciences and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

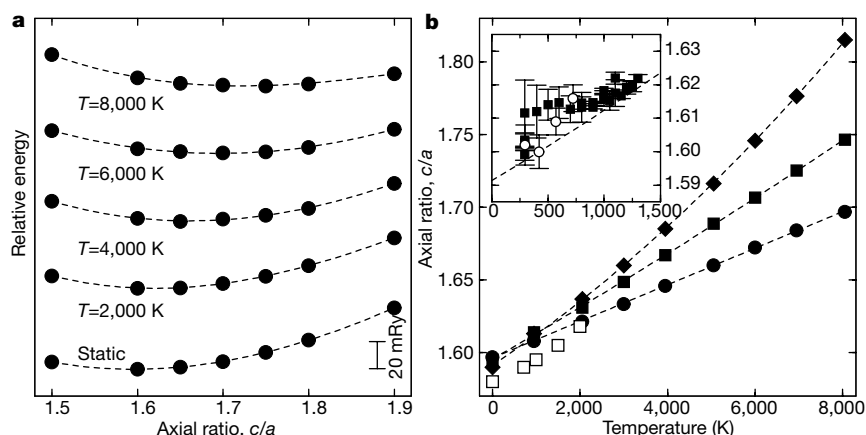
Seismological body-wave<sup>1</sup> and free-oscillation<sup>2</sup> studies of the Earth's solid inner core have revealed that compressional waves traverse the inner core faster along near-polar paths than in the equatorial plane. Studies have also documented local deviations from this first-order pattern of anisotropy on length scales ranging from 1 to 1,000 km (refs 3, 4). These observations, together with reports of the differential rotation<sup>5</sup> of the inner core, have generated considerable interest in the physical state and dynamics of the inner core, and in the structure and elasticity of its main constituent, iron, at appropriate conditions of pressure and temperature. Here we report first-principles calculations of the structure and elasticity of dense hexagonal close-packed (h.c.p.) iron at high temperatures. We find that the axial ratio  $c/a$  of h.c.p. iron increases substantially with increasing temperature, reaching a value of nearly 1.7 at a temperature of 5,700 K, where aggregate bulk and shear moduli match those of the inner core. As a consequence of the increasing  $c/a$  ratio, we have found that the single-crystal longitudinal anisotropy of h.c.p. iron at high temperature has the opposite sense from that at low temperature<sup>6,7</sup>. By combining our results with a simple model of polycrystalline texture in the inner core, in which basal planes are partially aligned with the rotation axis, we can account for seismological observations of inner-core anisotropy.

Although the inner core is probably not pure iron, the amount of light alloying element is small (2 to 3% mass fraction)<sup>8</sup> and may have a correspondingly small effect on elasticity<sup>8</sup>. The experimentally determined low-temperature, high-pressure phase of iron at least up to 300 GPa is h.c.p. ( $\epsilon$ -phase)<sup>9</sup>. This is also the liquidus phase up to at least 100 GPa (ref. 10), and there is theoretical evidence<sup>8,11</sup> that h.c.p. is the stable phase at inner-core conditions. Some experiments have been interpreted as indicating a different phase at high temperature and pressure<sup>12–14</sup>, but the evidence is inconclusive<sup>15</sup>, and the proposed structures<sup>13,14</sup> are closely related to h.c.p.

Despite the importance for our understanding of inner-core structure, the elastic-constant tensor of h.c.p. iron at the relevant pressure and temperature is currently unknown. Experimental measurements of the full elastic-constant tensor are so far restricted to ambient temperature. The isotropically averaged bulk and shear modulus at the pressure of the inner-core boundary and a temperature of 5,400 K has been reported<sup>16</sup>, but the full elastic-constant tensor at these conditions was not determined. The method of ref. 16, while promising, does not treat the electronic structure from first principles. Instead, an effective-potential model fit to short *ab initio* molecular dynamics trajectories is used. Calculations of the elastic constants with *ab initio* molecular dynamics may not currently be feasible, because long trajectories are required to obtain converged values from the relevant fluctuation formulas.

To overcome these limitations, we have developed a method that combines a first-principles treatment of the electronic structure





**Figure 1** Structure of h.c.p. iron at high pressure as a function of temperature. **a**, Relative energies ( $F$ ) for a density of  $13.04 \text{ Mg m}^{-3}$  as a function of axial ratio  $c/a$  for various temperatures. The scale bar refers to strain energy differences only. **b**, The axial ratio  $c/a$  as a function of temperature for three different densities (filled diamonds,  $12.52 \text{ Mg m}^{-3}$ ; filled squares,  $13.04 \text{ Mg m}^{-3}$ ; filled circles,  $13.62 \text{ Mg m}^{-3}$ ); dashed lines are quadratic fits

with an efficient model of the lattice vibrations that allows fully converged computations of the high-temperature elastic constants. Thermodynamic properties are governed by the Helmholtz free energy  $F$  as a function of volume ( $V$ ) and temperature ( $T$ )

$$F(V, T) = E_0(V) + E_{\text{el}}(V, T) - TS_{\text{el}}(V, T) + F_{\text{vib}}(V, T) \quad (1)$$

where  $E_0(V)$  is the static total energy,  $E_{\text{el}}(V, T)$  the contribution due to thermal excitations of the electrons,  $S_{\text{el}}(V, T)$  the electronic entropy, and  $F_{\text{vib}}(V, T)$  the phonon contribution. We obtain the first three terms on the right-hand side of equation (1) by an all-electron method (the linearized augmented plane wave method, LAPW), calculating  $E_{\text{band}} = E_0 + E_{\text{el}}$  directly and evaluating the Fermi–Dirac occupation of the electronic states for  $S_{\text{el}}$ .  $F_{\text{vib}}(V, T)$  is obtained using the particle-in-a-cell model<sup>17</sup>; this is a classical mean-field approach to lattice vibrations, in which one atom (the wanderer) is moved in the potential of the otherwise fixed lattice. The cell model includes intrinsic anharmonicity, which is important at the extreme conditions of the inner core. The factorized canonical partition function for  $F_{\text{vib}}$  is evaluated by computing the change in the electronic free energy  $F_{\text{el}}(V, T) = E_{\text{el}}(V, T) - TS_{\text{el}}(V, T)$  of a 48-atom supercell as a function of displacement of the wanderer along special symmetry directions<sup>17</sup>.

For efficient calculations of the energetics of the wanderer, we use a plane wave mixed basis method (PWMB). The use of a mixed basis, which contains localized basis functions in addition to plane waves, means that we are able to use harder pseudopotentials than are typical. Our Troullier–Martins pseudopotentials treat 3s, 3p and 3d states as valence states. Calculations based on harder

to the results. Open squares, previous theoretical work<sup>17</sup> at  $13.04 \text{ Mg m}^{-3}$ . Inset, comparison of our results at the lowest density (static pressure of 190 GPa, dashed line) with a polybaric set of experimental data at lower pressures. Open circles, data from ref. 19 (15–20 GPa); filled squares, data from ref. 20 (23–35 GPa).

pseudopotentials are expected to more nearly capture the all-electron limit: we reproduce pressures from the LAPW results at inner-core density to within 1%, and the elastic constants to within 2% r.m.s. The cell model has successfully been applied to calculations of thermodynamic properties of iron, including the Hugoniot up to inner-core densities<sup>17</sup>, as well as high-temperature thermodynamics of tantalum, including elastic constants<sup>18</sup>.

We obtain the equilibrium structure of h.c.p. iron over a range of densities and temperatures that span those of the inner core by computing  $F$  for a number of different values of  $c/a$  and determining the minimum free-energy value (Fig. 1a). We compute the full elastic-constant tensor by applying small magnitude finite strains to the lattice and evaluating the resulting changes in  $F(V, T)$ . By using volume-conserving strains, we assure the identity of the elastic constants calculated as thermodynamic variables with the stress-strain coefficients that are appropriate for elastic wave propagation<sup>7</sup>.

We find that the ratio  $c/a$  in h.c.p. iron changes considerably with temperature (Fig. 1): for a density of  $13.04 \text{ Mg m}^{-3}$ , this ratio increases from slightly less than the ideal value defined by hard-sphere packing (1.633) at low temperature to more than 1.7 at 6,000 K (Fig. 1b). This behaviour has been recognized before in experiments at lower pressure<sup>19,20</sup> and in calculations<sup>17</sup>, but is neglected in some recent theoretical work<sup>11,21</sup>.

The change in  $c/a$  with temperature is important, because it has a direct influence on the elastic anisotropy. As the  $c$ -axis expands with increasing temperature (at constant density), it becomes softer (more compressible). The corresponding longitudinal modulus,  $c_{33}$ , softens sufficiently that it becomes smaller than the other

**Table 1** Elastic moduli of dense h.c.p. iron at high temperatures

| $T$        | $C_{11}$<br>(GPa) | $C_{12}$<br>(GPa) | $C_{13}$<br>(GPa) | $C_{33}$<br>(GPa) | $C_{44}$<br>(GPa) | $C_{66}$<br>(GPa) | $K_S$<br>(GPa) | $\mu$<br>(GPa) | $\sigma$ |
|------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|----------------|----------------|----------|
| 0          | 1,880             | 880               | 770               | 2,065             | 450               | 500               | 1,185          | 505            | 0.31     |
| 4,000      | 2,035 (2,045)     | 1,435 (1,445)     | 810 (855)         | 1,650 (1,835)     | 270               | 300               | 1,325          | 335            | 0.38     |
| 5,000      | 2,050 (2,065)     | 1,720 (1,735)     | 860 (910)         | 1,555 (1,760)     | 210               | 165               | 1,375          | 240            | 0.42     |
| 6,000      | 2,130 (2,150)     | 2,010 (2,030)     | 925 (990)         | 1,470 (1,685)     | 140               | 60                | 1,425          | 145            | 0.45     |
| Inner core |                   |                   |                   |                   |                   |                   | 1,400          | 170            | 0.44     |

Elastic constants for h.c.p. iron calculated at a density of  $13.04 \text{ Mg m}^{-3}$  ( $V = 48 \text{ bohr}^3$  per atom) at various temperatures. We calculate the Helmholtz free energy (equation (1)) by combining LAPW results for the electronic terms and a particle-in-a-cell model for the phonon contribution (see text). Computational parameters for the LAPW calculations are given in ref. 7. For the particle-in-a-cell method we evaluate the energetics of a wanderer atom in a 48-atom supercell with a Brillouin zone sampling of 8  $k$ -points. To efficiently perform calculations for the wanderer displacement, we use a PWMB method with energy cut-offs of 1,400 eV and 80 eV in the muffin-tin spheres and the interstitial region, respectively. In the PWMB method we use Troullier–Martins pseudopotentials which are constructed using 3s, 3p and 3d valence states (3d local) with cut-off radii of 1.5 bohr to avoid pseudopotential overlap at the small unit cell volume of this study.  $C_{66} = 1/2(C_{11} - C_{12})$  is a shear elastic constant and added for comparison with  $C_{44}$ . Isothermal elastic constants are converted to adiabatic constants (shown in parentheses) using thermodynamic derivatives of  $F(V, T)$ . They are calculated self-consistently from our results. Adiabatic aggregate moduli  $K_S$ ,  $\mu$  and Poisson's ratio ( $\sigma$ ) are given for comparison with seismic properties of the inner core at the same density (ref. 24).

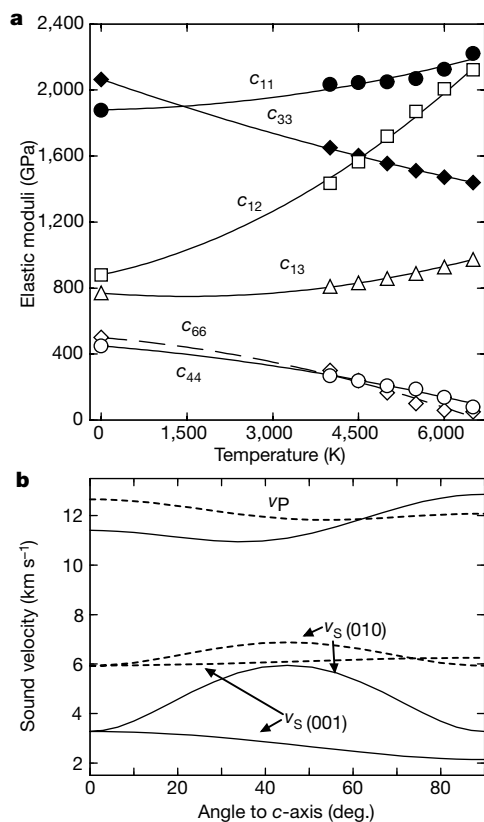
longitudinal modulus  $c_{11}$  (Fig. 2a). As a result, the sense of the longitudinal anisotropy is reversed at high temperature, with compressional acoustic wave propagation being faster in the basal plane than along the  $c$  axis (Fig. 2b). The off-diagonal elastic constants are also affected by the temperature-induced change in structure:  $c_{12}$  increases rapidly with temperature, because the basal plane shrinks with increasing temperature at constant density. The shear constants  $c_{44}$  and  $c_{66} = 1/2(c_{11} - c_{12})$  show a strong temperature dependence, and decrease by a factor of almost four at 6,000 K and change order as well (Fig. 2a). As a result, the velocity of shear waves is considerably smaller at high temperature and the sense of shear anisotropy is reversed (Fig. 2b), with the propagation of the (001) polarized shear wave becoming faster along the  $c$ -axis than in the basal plane. In agreement with ref. 16, our calculations imply a shear instability in h.c.p. iron at very high temperature ( $c_{66}$  approaches zero at 7,000 K).

The temperature-induced reversal in longitudinal anisotropy that we find would not have been predicted on the basis of the behaviour of other h.c.p. transition metals. At ambient pressure, many h.c.p. transition metals show little change in the ratio  $c_{11}/c_{33}$  with temperature: titanium shows the largest such change, with a decrease of 13% up to a temperature of 1,075 K, about half its melting point. Our results suggest that absolute temperature rather than the temperature relative to melting controls anisotropy. The character-

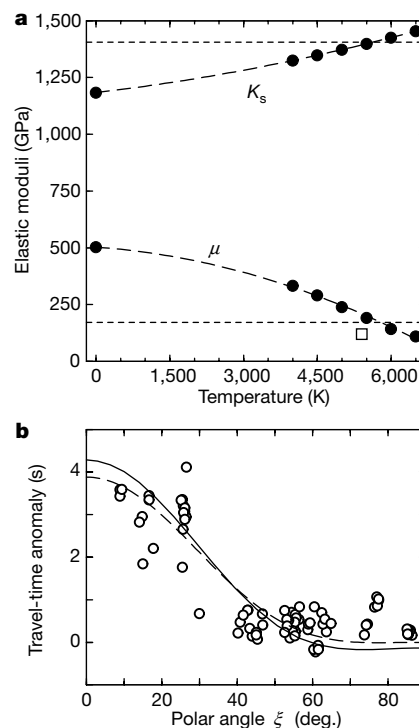
istics of elasticity we find at high temperature invalidate previous attempts to explain inner-core anisotropy using low-temperature elastic constants of iron<sup>6,22,23</sup>.

To compare our results with seismological observations, we convert the computed isothermal elastic constants to adiabatic elastic constants, and from these we calculate the adiabatic bulk modulus ( $K_S$ ) and shear modulus ( $\mu$ ) using Hashin–Shtrikman bounds (Fig. 3a, Table 1). The aggregate moduli we determine agree with those of the inner core at a temperature of 5,700 K. Poisson's ratio at 5,700 K (0.44) is in excellent agreement with previous theoretical estimates<sup>16</sup>, and with the seismologically determined value for the inner core<sup>24</sup>. This shows that the high Poisson's ratio of the inner core can be accounted for by the effects of temperature and pressure on the elasticity of iron alone; it is unnecessary to invoke shear attenuation and dispersion or the presence of partial melt in the inner core, as has been suggested<sup>22</sup>.

As the inner core is nearly isothermal<sup>8</sup>, the comparison of  $K_S$  and  $\mu$  to those determined seismologically for the inner core provides a way to estimate the temperature in the Earth's deep interior. This approach is complementary to estimates based on the melting temperature of iron, which suffer uncertainty due to the unknown but possibly large influence of light alloying elements<sup>15,16,21</sup>. Assuming a melting point depression of a few hundred degrees<sup>15</sup>, our value for inner-core temperature of 5,700 K is consistent with estimates of the melting point of pure iron<sup>21</sup> (6,400 K) and of the Hugoniot in shock wave experiments (5,000 to 5,700 K at 245 GPa)<sup>12</sup>. The theoretical result of ref. 16 (5,400 K) and the extrapolated melting curve from static experiments<sup>15</sup> (~5,000 K) are both somewhat lower.



**Figure 2** Elasticity of h.c.p. iron at a density of  $13.04 \text{ Mg m}^{-3}$ . **a**, Single-crystal elastic constants as a function of temperature. Static values are connected to the high-temperature results with a quadratic fit (solid lines). Data are shown for the longitudinal elastic constants  $c_{11}$  (filled circles) and  $c_{33}$  (filled diamonds), the off-diagonal constants  $c_{12}$  (squares) and  $c_{13}$  (triangles), and the shear elastic constants  $c_{44}$  (circles) and  $c_{66}$  (diamonds).  $c_{66} = 1/2(c_{11} - c_{12})$  (dashed line) is not an independent elastic constant but is included for comparison with  $c_{44}$ . **b**, Acoustic anisotropy in the single crystal: wave velocities for the compressional wave ( $v_p$ ) and the two polarizations of the shear wave ( $v_s$ , polarization direction indicated in brackets) are shown as a function of propagation direction with respect to the  $c$ -axis. Results at 6,000 K (solid lines) are compared to static results (dashed lines).



**Figure 3** Acoustic properties of iron and the inner core. **a**, Comparison of iron's adiabatic bulk modulus ( $K_S$ ) and shear modulus ( $\mu$ ) as a function of temperature (filled symbols, long-dashed lines) with the values of the inner core at the same density ( $13.04 \text{ Mg m}^{-3}$ ) (short-dashed lines, ref. 24). For comparison a previous theoretical result (open square, ref. 16) is included. **b**, The differential travel-time anomaly between BC and DF branches of the PKP core seismic phases due to inner-core anisotropy as a function of propagation direction. Our result (solid line) is for a model of the inner core in which 1/3 of the crystals have their basal planes aligned with the rotation axis. The observations of ref. 1 (open circles and dashed line) are shown for comparison. We convert from anisotropy in P-wave velocity for our model to differential travel time as outlined in ref. 6.

The sense of anisotropy that we find at high temperature changes our view of the polycrystalline texture of the inner core and the dynamic processes that may produce it. On the basis of our results, we propose a simple model of the polycrystalline texture of the inner core that explains the main features of its anisotropy. We find that if one-third of the basal planes are aligned with the Earth's rotation axis in an otherwise random medium, compressional wave travel-time anomalies are well explained (Fig. 3b). This model is almost certainly over-simplified. The key element is the tendency for the fast crystallographic direction (*a*) to be aligned with the observed symmetry axis of inner-core anisotropy (approximately polar). It is probable that the actual direction and degree of crystallographic alignment will vary with geographic location. Such variations may account for seismological observations of heterogeneity<sup>3,4</sup>.

Important remaining questions include the origin of polycrystalline texture in the inner core, which may have been acquired during solidification of the inner core<sup>25</sup> or may have developed subsequently as a result of plastic deformation. If plastic deformation is the prevalent mechanism, the texture must depend on the dominant microscopic deformation mechanisms in iron at inner-core conditions and the source of stress that leads to the preferred alignment. While these are currently unknown, our simple textural model is consistent with the available information.

Candidates for the deformation mechanism include diffusion<sup>23</sup> and dislocation glide<sup>26,27</sup>. Among the crystallographic slip planes that may participate in dislocation glide, the basal plane is the easiest, according to recent experimental work<sup>27</sup>. The temperature-induced increase in *c/a* that we find may make basal slip even more favourable at high temperature. Several sources of the external stress field have been proposed: it may arise from aspherical growth of the inner core<sup>23</sup>, gravitational coupling with the Earth's mantle<sup>28</sup>, thermal convection in the inner core<sup>26</sup>, or Maxwell stresses created by the Earth's magnetic field<sup>29,30</sup>. The pattern of flow in each of these cases is distinct; many models, however, share a polar-dominated flow, that is, a pattern of flow in which the highest velocities are oriented parallel to the polar axis.

If dislocation glide is the most important deformation mechanism, the easiest glide plane would be preferentially aligned with the dominant direction of flow, parallel to the rotation axis of the Earth. If diffusion dominates, preferred alignment is driven by minimization of strain energy, in which case the elastically stiffest (fastest) direction is aligned with the rotation axis<sup>23</sup>. In either case, basal planes would tend to be aligned with the pole, consistent with our simple textural model. □

Received 4 April; accepted 13 July 2001.

1. Song, X. D. & Helmberger, D. V. Anisotropy of the Earth's inner core. *Geophys. Res. Lett.* **20**, 2591–2594 (1993).
2. Tromp, J. Support for anisotropy of the Earth's inner core from free oscillations. *Nature* **366**, 678–681 (1993).
3. Vidale, J. E. & Earle, P. S. Fine scale heterogeneity in the Earth's inner core. *Nature* **404**, 273–275 (2000).
4. Creager, K. C. Large-scale variations in inner core anisotropy. *J. Geophys. Res.* **104**, 23127–23139 (1999).
5. Song, X. D. & Richards, P. G. Seismological evidence for differential rotation of the Earth's inner core. *Nature* **382**, 221–224 (1996).
6. Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth's inner core. *Science* **267**, 1972–1975 (1995).
7. Steinle-Neumann, G., Stixrude, L. & Cohen, R. E. First-principles elastic constants for the hcp transition metals Fe, Co, and Ni at high pressure. *Phys. Rev. B* **60**, 791–799 (1999).
8. Stixrude, L., Wasserman, E. & Cohen, R. E. Composition and temperature of the Earth's inner core. *J. Geophys. Res.* **102**, 24729–24739 (1997).
9. Mao, H.-K., Wu, Y., Chen, Y., Shu, J. F. & Jephcoat, A. P. Static compression of iron to 300 GPa and Fe<sub>0.8</sub>Ni<sub>0.2</sub> alloy to 260 GPa: implications for composition of the core. *J. Geophys. Res.* **95**, 21737–21742 (1990).
10. Shen, G., Mao, H.-K., Hemley, R. J., Duffy, T. S. & Rivers, M. L. Melting and crystal structure of iron at high pressures and temperatures. *Geophys. Res. Lett.* **25**, 373–376 (1998).
11. Vočadlo, L. *et al.* Ab initio free energy calculations on the polymorphs of iron at core conditions. *Phys. Earth Planet. Inter.* **117**, 123–137 (2000).
12. Brown, J. M. & McQueen, R. Phase transitions, Grüneisen parameter, and elasticity for shocked iron between 77 GPa and 400 GPa. *J. Geophys. Res.* **91**, 7485–7494 (1986).
13. Dubrovinsky, L. S. *et al.* In situ X-ray study of thermal expansion and phase transition of iron at multimegabar pressure. *Phys. Rev. Lett.* **84**, 1720–1724 (2000).

14. Andrault, D., Fiquet, G., Kunz, M., Viscoekas, F. & Häusermann, D. The orthorhombic structure of iron: an in situ study at high-temperature and high-pressure. *Science* **278**, 831–834 (1997).
15. Boehler, R. High-pressure experiments and the phase diagram of lower mantle and core materials. *Rev. Geophys.* **38**, 221–245 (2000).
16. Laio, A., Bernard, S., Chiarotti, G. L., Scandolo, S. & Tosatti, E. Physics of iron at Earth's core condition. *Science* **287**, 1027–1030 (2000).
17. Wasserman, E., Stixrude, L. & Cohen, R. E. Thermal properties of iron at high pressures and temperatures. *Phys. Rev. B* **53**, 8296–8309 (1996).
18. Gülseren, O., Cohen, R. E. & Krakauer, H. Equation of state and elasticity of Fe at high pressure. *Bull. Am. Phys. Soc.* **44**, 1490 (1999).
19. Huang, E., Bassett, W. A. & Tao, P. Pressure-temperature-volume relationship for hexagonal close packed iron determined by synchrotron radiation. *J. Geophys. Res.* **92**, 8129–8135 (1987).
20. Funamori, N., Yagi, T. & Uchida, T. High-pressure and high-temperature in situ x-ray diffraction study of iron to above 30 GPa using MA8-type apparatus. *Geophys. Res. Lett.* **23**, 953–956 (1996).
21. Alfè, D., Gillan, M. J. & Price, G. D. The melting curve of iron at pressures of the Earth's core from *ab initio* calculations. *Nature* **401**, 462–464 (1999).
22. Singh, S. C., Taylor, M. A. J. & Montagner, J. P. On the presence of liquid in Earth's inner core. *Science* **287**, 2471–2474 (2000).
23. Yoshida, S., Sumita, I. & Kumazawa, M. Growth model of the inner core coupled with outer core dynamics and the resultant elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103 (1996).
24. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
25. Bergman, M. I. Measurements of elastic anisotropy due to solidification texturing and the implications for the Earth's inner core. *Nature* **389**, 60–63 (1997).
26. Jeanloz, R. & Wenk, H.-R. Convection and anisotropy of the inner core. *Geophys. Res. Lett.* **15**, 72–75 (1988).
27. Wenk, H.-R., Matthies, S., Hemley, R. J., Mao, H.-K. & Shu, J. The plastic deformation of iron at pressures of the Earth's inner core. *Nature* **405**, 1044–1046 (2000).
28. Buffett, B. A. Geodynamic estimates of the viscosity of the Earth's inner core. *Nature* **388**, 571–573 (1997).
29. Karato, S.-I. Seismic anisotropy of the Earth's inner core resulting from flow induced by Maxwell stresses. *Nature* **402**, 871–873 (1999).
30. Buffett, B. A. & Wenk, H.-R. Texturing of the Earth's inner core by Maxwell stresses. *Nature* **413**, 60–63 (2001).

## Acknowledgements

We thank B. Buffett and B. Kiefer for discussions, and H. Krakauer and D. Singh for the use of their LAPW code. This work was supported by the US National Science Foundation (R.E.C. and L.S.) and the US Department of Energy (R.E.C.). Calculations were performed on the CRAY SV1 at the Geophysical Laboratory, supported by the US National Science Foundation and the W. M. Keck Foundation.

Correspondence and requests for materials should be addressed to G.S.-N. (e-mail: gerd@umich.edu).

## Texturing of the Earth's inner core by Maxwell stresses

B. A. Buffett\* & H.-R. Wenk†

\* Department of Earth & Ocean Sciences, University of British Columbia, Vancouver, V6T 1Z4 Canada

† Department of Earth & Planetary Science, University of California, Berkeley, California 94720, USA

Elastic anisotropy in the Earth's inner core has been attributed to a preferred lattice orientation<sup>1</sup>, which may be acquired during solidification of the inner core<sup>2</sup> or developed subsequent to solidification as a result of plastic deformation<sup>3–5</sup>. But solidification texturing alone cannot explain the observed depth dependence of anisotropy<sup>6–8</sup>, and previous suggestions for possible deformation processes have all relied on radial flow, which is inhibited by thermal<sup>9</sup> and chemical stratification<sup>10</sup>. Here we investigate the development of anisotropy as the inner core deforms plastically under the influence of electromagnetic (Maxwell) shear stresses. We estimate the flow caused by a representative magnetic field using polycrystal plasticity simulations for  $\epsilon$ -iron, where the imposed deformation is accommodated by basal and prismatic slip<sup>11</sup>. We find that individual grains in an initially random polycrystal become preferentially oriented with their



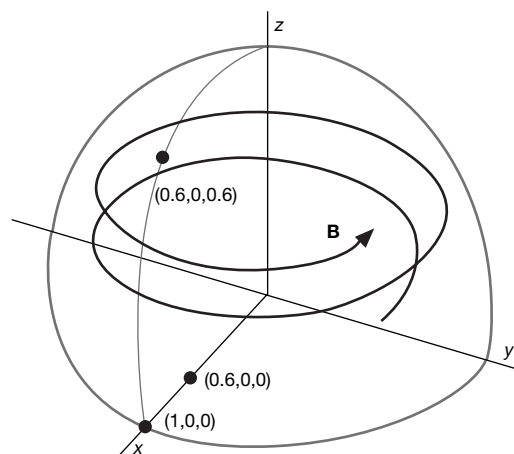
*c* axes parallel to the equatorial plane. This pattern is accentuated if deformation is accompanied by recrystallization. Using the single-crystal elastic properties of  $\epsilon$ -iron at core pressure and temperature<sup>12</sup>, we average over the simulated orientation distribution to obtain a pattern of elastic anisotropy which is similar to that observed seismologically<sup>13,14</sup>.

Observational evidence for elastic anisotropy in the interior of the inner core is derived from travel-time measurements of body waves<sup>13</sup> and anomalous splitting of core-sensitive free oscillations<sup>14</sup>. Both types of observations suggest that compressional waves travel faster along polar paths than equatorial paths. There is also general agreement that the inner core is composed of an iron-rich alloy, which forms a hexagonal close-packed (h.c.p.) structure<sup>15</sup> (known as  $\epsilon$ -iron). However, there is much less consensus on the single-crystal properties of iron at high pressures and temperatures. First-principles calculations<sup>16</sup> for 0 K and a fixed density of 13 Mg m<sup>-3</sup> predict that  $\epsilon$ -iron single crystals are only weakly anisotropic, whereas experiments at high pressure<sup>17</sup> suggest much stronger anisotropy. More recent calculations that allow for finite temperature<sup>12</sup> predict 10–12% elastic anisotropy at inner-core pressure and temperature, with the fast direction perpendicular to the *c* axis. According to these new calculations, the observed anisotropy can be explained by preferentially aligning the *c* axes of the h.c.p. crystals perpendicular to the rotation axis of the inner core. Such an alignment can be produced by Maxwell shear stresses.

Geodynamo models<sup>18</sup> provide a plausible description of the magnetic field in the vicinity of the inner core. The predicted fields typically comprise a strong azimuthal component  $B_\phi$  that wraps around the inner core, and a vertical component  $B_z$  that is aligned with the rotation axis. The combination of these features produces an electromagnetic force with two orthogonal components. A radial component is directed into the inner core, and a tangential component is aligned with the azimuthal coordinate  $\phi$  (see Fig. 1) The radial component of the force depends only on the  $B_\phi$  component of the field<sup>5</sup>, and produces relatively little flow when the inner core is stratified<sup>19</sup>. We focus on the tangential force, which can be characterized by a Maxwell shear stress  $B_\phi B_\mu / \mu$ , where  $\mu$  is the magnetic permeability of free space. For a representative magnetic field, we take  $B_z$  to be spatially constant and let

$$B_\phi = B_\phi(a) \left( \frac{r}{a} \right)^2 \cos \theta \sin \theta \quad (1)$$

where  $B_\phi(a)$  is a constant,  $r$  is radius,  $\theta$  is co-latitude and



**Figure 1** Schematic illustration of the inner core and the magnetic field **B**. Polycrystal plasticity simulations were performed for three locations in the inner core, which are identified by cartesian coordinates ( $x/a, y/a, z/a$ ), where  $a = 1,221$  km is the present-day radius of the inner core.

$a = 1,221$  km is the present-day radius of the inner core. The tangential component of the electromagnetic force  $\mathbf{F} = (\nabla \times \mathbf{B}) \times \mathbf{B} / \mu$  is given by

$$F_\phi = \left( \frac{B_z B_\phi(a) r}{\mu a^2} \right) \sin \theta \quad (2)$$

For  $B_z = -2$  mT and  $B_\phi(a) = 20$  mT (ref. 18), we obtain a magnetic shear stress of the order of several pascals.

The electromagnetic force in equation (2) must be balanced by an opposing force to prevent unbounded acceleration of the inner core. Gravitational forces are probably sufficient to balance  $F_\phi$  and thereby maintain a steady rate of inner-core rotation. We also expect gravitational forces to deform the inner core, because the surface of the inner core continually relaxes toward a constant potential surface. However, calculations show that the flow is periodic in longitude<sup>20</sup>, so that the strain due to the gravitational forces averages to zero as the inner core rotates. By contrast,  $F_\phi$  causes a steady flow and the associated strain increases with time. The torque due to the gravitational forces is approximated in the deformation calculation as an equivalent shear stress on the spherical surface immediately below the boundary topography, and  $F_\phi$  is applied as a body force in the interior. The flow induced in a uniform viscous sphere, neglecting an arbitrary rigid-body rotation, is given by

$$v_\phi = - \frac{1}{10} \frac{B_z B_\phi(a) r^3}{\mu \eta a^2} \sin \theta \quad (3)$$

where  $\eta$  is the effective viscosity of the inner core. The velocity gradient tensor calculated from equation (3) is used in polycrystal plasticity simulations to predict the reorientation of a random aggregate of h.c.p. crystals.

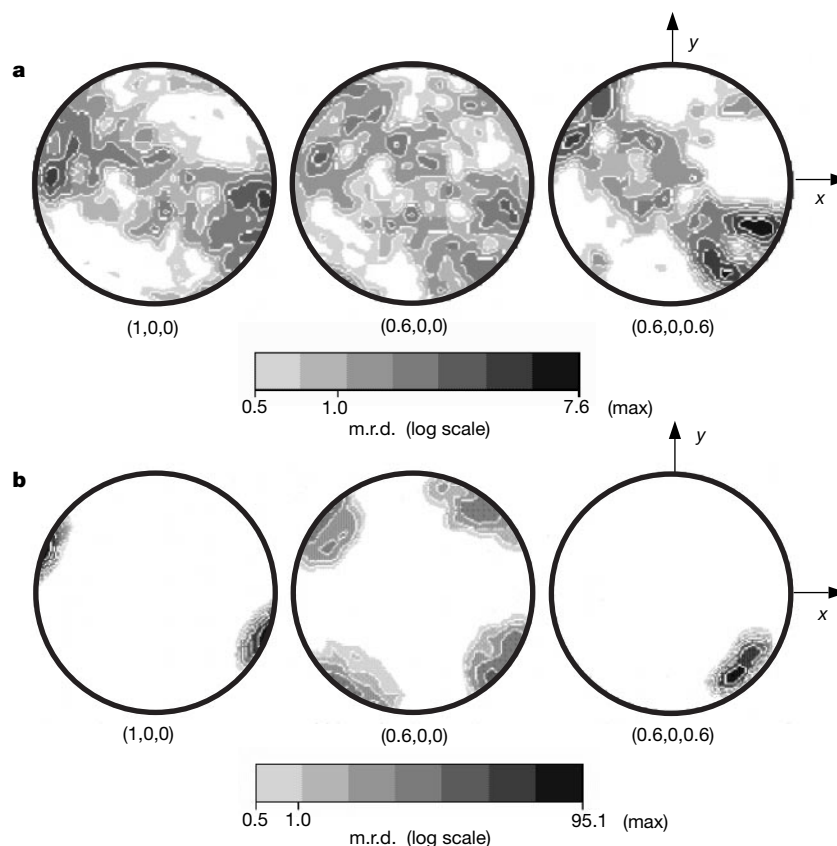
A polycrystal deforms while maintaining local stress equilibrium and strain compatibility across grain boundaries. This is achieved by allowing variations in stress and strain between grains and within individual grains. Modelling such local plastic heterogeneity is difficult, and most simulations use simplifications. Differences between common modelling assumptions constitute only a second-order effect compared with larger uncertainties in the physical properties of the inner core. We apply a viscoplastic self-consistent model<sup>21</sup> that permits grains to deform individually, depending on orientation. This model, in its one-site approximation<sup>22</sup>, regards each grain of the polycrystal as a viscoplastic inclusion in an effective medium with the average viscoplastic properties of the polycrystal. In general, the effective medium is anisotropic and evolves with time.

The model requires as input the initial crystallographic orientations of the constituent grains and the active intracrystalline deformation mechanisms. An aggregate of 500 grains with random crystallographic orientation serves as the starting texture in our simulations. Potential slip systems and their critical resolved shear stress ratios are listed in Table 1. Deformation experiments in diamond anvil cells<sup>11</sup> indicate that basal slip is active in  $\epsilon$ -iron at 220 GPa. We assume that other slip systems commonly observed in h.c.p. metals are active as well, but that twinning is absent. Previous simulations<sup>3,11</sup> indicate that the development of texture is insensitive to the relative ease of basal and prismatic slip, so we assume that

**Table 1** Slip systems of h.c.p. metals used in our simulations

| Description | Plane            | Direction              | Critical resolved shear stress ratio |
|-------------|------------------|------------------------|--------------------------------------|
| Basal       | (0001)           | ( $\bar{1}2\bar{1}0$ ) | 1                                    |
| Prismatic   | {10 $\bar{1}0$ } | ( $\bar{1}2\bar{1}0$ ) | 1                                    |
| *Pyramidal  | {10 $\bar{1}1$ } | ( $\bar{1}2\bar{1}0$ ) | 10                                   |
| *Pyramidal  | {10 $\bar{1}1$ } | (11 $\bar{2}$ 3)       | 10                                   |

\*Included but not significantly active.



**Figure 2** Simulated texture development in  $\epsilon$ -iron. The deformation is imposed in equal increments for a total strain of roughly 40%. Orientation distributions for the  $c$  axis at three positions in the inner core are shown for the cases of pure deformation (**a**) and deformation with recrystallization (**b**). Each orientation distribution is expressed in

multiples of a random distribution (m.r.d.) and the plots are oriented such that the north pole is in the centre and the equator is at the periphery. Recrystallization produces stronger textures, but does not change the orientation of the anisotropy.

both slip systems operate with equal ease. Additional parameters in the calculations<sup>22</sup> include the stress exponent ( $n = 3$ ) in the non-linear viscous rheology, and the hardening coefficient ( $h = 0.1$ ), which produces only minor hardening.

Recrystallization of a polycrystal may modify the texture that develops through plastic deformation, especially at high temperature. Modifications to include the effects of dynamic and static recrystallization in the simulations<sup>23</sup> are based on a competition between growth of less deformed grains and nucleation of new domains in highly deformed grains. The driving force for recrystallization in the model is the strain energy accumulated during deformation. In our simulations, grain growth is favoured over nucleation to reflect the effects of high temperature and low deviatoric stresses in the inner core.

Incremental changes in strain are imposed in the polycrystal plasticity simulations using the steady flow given in equation (3). Accumulating 40 equal increments yields a total strain of about 40% near the surface of the inner core. The duration of each step is controlled by our choice of effective viscosity  $\eta$ . For values between  $10^{14}$  and  $10^{16}$  Pa s, the simulation requires roughly 1–100 Myr to develop the 40% strain. When recrystallization is included in the simulation, grains are allowed to grow or shrink according to their relative strain energy after every deformation step.

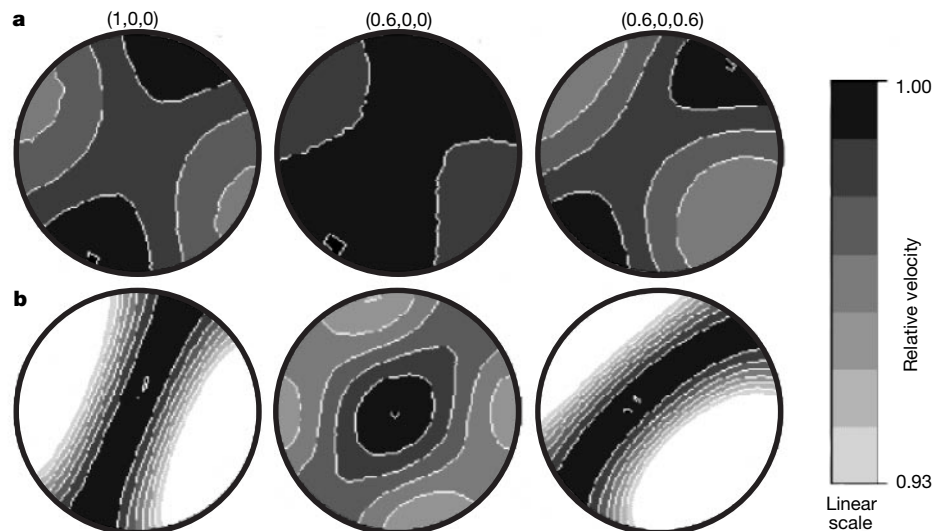
Continuous orientation distributions are calculated from the individual grain orientations in the simulation<sup>24</sup>. Figure 2a shows the orientation distributions for the  $c$  axis after 40 deformation steps (with no recrystallization). These plots are not projections of the Earth but illustrate orientation distributions at three different locations in the inner core. Each plot is oriented relative to Earth coordinates such that the centre is parallel to the polar axis and the

periphery is parallel to the equatorial plane. The  $x$  axis in the equatorial plane of the inner core is directed to the right and the  $y$  axis is directed towards the top of each plot.

Position (1,0,0) is on the equatorial plane at the surface of the inner core. The  $c$  axes at this location are preferentially aligned in the equatorial plane and nearly parallel to the  $x$  direction. The degree of alignment is expressed in multiples of a random distribution (m.r.d.). Thus the peak pole density in Fig. 2a is nearly eight times a random distribution. The polycrystal at (1,0,0) is deformed in simple shear, so we infer that the motion after 40 deformation steps is accommodated mainly by basal slip. Similar orientation patterns are predicted at locations (0.6,0,0) and (0.6,0,0.6). The intensity of alignment at (0.6,0,0) is weaker because the electromagnetic force decreases into the inner core. Position (0.6,0,0.6) at mid-latitudes is closer to the surface, and the intensity of alignment increases.

Including the effect of dynamic recrystallization in the simulations allows less-deformed grains to grow at the expense of the rest. The least-deformed grains in the simulation have  $c$  axes aligned in the equatorial plane, whereas the most-deformed grains have  $c$  axes parallel to the polar axis. As the deformed grains vanish during recrystallization, the orientation distribution sharpens considerably. The peak orientation distribution in Fig. 2 is nearly 100 times a random distribution.

Aggregate elastic properties at the three locations are calculated using the adiabatic single-crystal elastic constants of ref. 4 (at 6,000 K) and the orientation distributions predicted in Fig. 2. We use a geometric average, which is intermediate between the upper and lower bounds. Figure 3 shows the predicted variations in P-wave velocity  $v_p$  as a function of direction. Only relative variations



**Figure 3** Averaged aggregate P-wave velocity for  $\epsilon$ -iron polycrystals. Calculations at positions (1,0,0), (0.6,0,0) and (0.6,0,0.6) are based on single-crystal elastic properties<sup>12</sup> and the orientation distributions predicted in Fig. 2. We show the relative variation in

P-wave velocity for the cases of pure deformation (**a**) and deformation with recrystallization (**b**). The aggregate elastic properties are more sensitive to texture pattern than to texture strength.

are plotted, by assigning the fastest  $v_p$  in each panel a value of 1.0. The slowest directions are in the equatorial plane (that is, perpendicular to  $z$ ). The difference between the fastest and slowest direction is about 3% without recrystallization (Fig. 3a). Including recrystallization increases the anisotropy to about 11% (Fig. 3b), close to the single-crystal value.

To assess the compatibility of our predictions with seismological observations, we adopt an elastic model in which  $v_p$  in the radial direction is 3% slower than  $v_p$  in the perpendicular directions. Values of  $v_p$  for intermediate directions are inferred from Fig. 3a and vary approximately as  $\cos 2\theta$ , where  $\theta$  is the angle from the slow direction. We then integrate for the travel time of P-waves on polar and equatorial paths through the inner core (assuming straight ray-paths). The difference in travel times for equatorial and polar paths is 2.9 to 5.0 seconds for ray-paths that penetrate 500 to 700 km into the inner core. These travel-time differences have correct sign and sufficient amplitude to explain the seismological observations.

Our representative description of the magnetic field inside the inner core predicts that both strain and texture will be at their maximum values near the surface of the inner core. However, as the inner core grows, this texture will gradually become buried. The apparent absence of anisotropy from the uppermost few hundred kilometres of the inner core<sup>6–8</sup> may indicate that this region has had insufficient time to develop strains of the order of 50–100%. Relative rotation between the inner and outer core would tend to average the magnetic stresses and produce texture with a cylindrical structure, while persistent complexities in the form of the magnetic field could cause departures from this simple structure<sup>25,26</sup>. Further complications could arise from an initial solidification texture, partial melt<sup>27</sup>, lateral temperature variations or flow in the surrounding liquid outer core. Each of these additional effects warrant quantitative assessment. □

Received 12 January; accepted 18 July 2001.

1. Song, X. Anisotropy of the Earth's inner core. *Rev. Geophys.* **35**, 297–313 (1997).
2. Bergman, M. I. Measurements of elastic anisotropy due to solidification texturing and the implications for the Earth's inner core. *Nature* **389**, 60–63 (1997).
3. Wenk, H.-R., Baumgardner, J. R., Lebensohn, R. A. & Tomé, C. N. A convective model to explain anisotropy of the inner core. *J. Geophys. Res.* **105**, 5663–5677 (2000).
4. Yoshida, S., Sumita, I. & Kumazawa, M. Growth model of the inner core coupled with the outer core dynamics and the resulting elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103 (1996).
5. Karato, S. Seismic anisotropy of the Earth's inner core resulting from flow induced by Maxwell stresses. *Nature* **402**, 871–873 (1999).

6. Song, X. & Helmberger, D. V. Depth dependence of anisotropy in Earth's inner core. *J. Geophys. Res.* **100**, 9805–9816 (1995).
7. Creager, K. C. Large-scale variations in inner core anisotropy. *J. Geophys. Res.* **104**, 23127–23139 (1999).
8. Garcia, R. & Souriau, A. Inner core anisotropy and heterogeneity level. *Geophys. Res. Lett.* **27**, 3121–3124 (2000).
9. Yukutake, T. Implausibility of thermal convection in the Earth's solid inner core. *Phys. Earth Planet. Inter.* **101**, 1–13 (1998).
10. Stacey, F. D. Theory of thermal and elastic properties of the lower mantle and core. *Phys. Earth Planet. Inter.* **89**, 219–245 (1995).
11. Wenk, H.-R., Matthies, S., Hemley, R. J., Mao, H.-K. & Shu, J. The plastic deformation of iron at pressures of the Earth's inner core. *Nature* **405**, 1044–1046 (2000).
12. Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Gülseren, O. Elasticity of iron at the temperature of the Earth's inner core. *Nature* **413**, 57–60 (2001).
13. Creager, K. C. Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP. *Nature* **356**, 309–314 (1992).
14. Tromp, J. Support for anisotropy of the Earth's inner core from free oscillations. *Nature* **366**, 678–681 (1993).
15. Yoo, C. S., Akella, J., Campbell, A. J. & Mao, H.-K. Phase diagram of iron by in situ x-ray diffraction: Implications for Earth's core. *Science* **270**, 1473–1475 (1995).
16. Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth's inner core. *Science* **267**, 1972–1975 (1995).
17. Mao, H. K. *et al.* Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core. *Nature* **396**, 741–743 (1998).
18. Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth's inner core. *Science* **274**, 1887–1891 (1996).
19. Buffett, B. A. & Bloxham, J. Deformation of Earth's inner core by electromagnetic forces. *Geophys. Res. Lett.* **27**, 4001–4004 (2000).
20. Buffett, B. A. Geodynamic estimates of the viscosity of the Earth's inner core. *Nature* **388**, 571–573 (1997).
21. Molinari, A., Canova, G. R. & Ahzi, S. A self-consistent approach of the large deformation polycrystal viscoplasticity. *Acta Metall.* **35**, 2983–2994 (1987).
22. Lebensohn, R. A., Wenk, H.-R. & Tomé, C. N. Modelling deformation and recrystallization textures in calcite. *Acta Mater.* **46**, 2683–2693 (1998).
23. Wenk, H.-R. & Tomé, C. N. Modeling dynamic recrystallization of olivine deformed in simple shear. *J. Geophys. Res.* **104**, 25513–25527 (1999).
24. Wenk, H.-R., Matthies, S., Donovan, J. & Chateigner, D. BEARTEX, a Windows-based program system for quantitative texture analysis. *J. Appl. Crystallogr.* **31**, 262–269 (1998).
25. Tanaka, S. & Hamaguchi, H. Degree one heterogeneity and hemispherical variation of anisotropy in the inner core from PKP(BC) - PKP(DF) times. *J. Geophys. Res.* **102**, 2925–2938 (1997).
26. Vidale, J. S. & Earle, P. S. Fine-scale heterogeneity in the Earth's inner core. *Nature* **404**, 273–275 (2000).
27. Singh, S. C., Taylor, M. A. J. & Montagner, J. P. On the presence of liquid in Earth's inner core. *Science* **287**, 2471–2474 (2000).

## Acknowledgements

We thank G. Steinle-Neumann for sending us a preprint of his work. B.A.B. was supported by NSERC, and H.R.W. acknowledges support from IGPP-LANL and NSF.

Correspondence and requests for materials should be addressed to B.A.B. (e-mail: buffett@geop.ubc.ca).



# Human presence in the European Arctic nearly 40,000 years ago

Pavel Pavlov<sup>\*†</sup>, John Inge Svendsen<sup>†‡</sup> & Svein Indrelid<sup>§</sup>

<sup>\*</sup> Institute of Language, Literature and History, Komi Scientific Center, Russian Academy of Sciences, Ural Division, Kommunisticheskaya st. 26, 167000, Syktyvkar, Komi Republic, Russia

<sup>‡</sup> Centre for Studies of the Environment and Resources, University of Bergen, Høyteknologisenteret (HIB), N-5020 Bergen, Norway

<sup>§</sup> Bergen Museum, University of Bergen, Harald Hårfagresgt.1, N-5020 Bergen, Norway

<sup>†</sup> These authors contributed equally to the work

The transition from the Middle to the Upper Palaeolithic, approximately 40,000–35,000 radiocarbon years ago, marks a turning point in the history of human evolution in Europe. Many changes in the archaeological and fossil record at this time have been associated with the appearance of anatomically modern humans<sup>1,2</sup>. Before this transition, the Neanderthals roamed the continent, but their remains have not been found in the northernmost part of Eurasia. It is generally believed that this vast region was not colonized by humans until the final stage of the last Ice Age some 13,000–14,000 years ago<sup>3,4</sup>. Here we report the discovery of traces of human occupation nearly 40,000 years old at Mamontovaya Kurya, a Palaeolithic site situated in the European part of the Russian Arctic. At this site we have uncovered stone artefacts, animal bones and a mammoth tusk with human-made marks from strata covered by thick Quaternary deposits. This is the oldest documented evidence for human presence at this high latitude; it implies that either the Neanderthals expanded much further north than previously thought or that modern humans were present in the Arctic only a few thousand years after their first appearance in Europe.

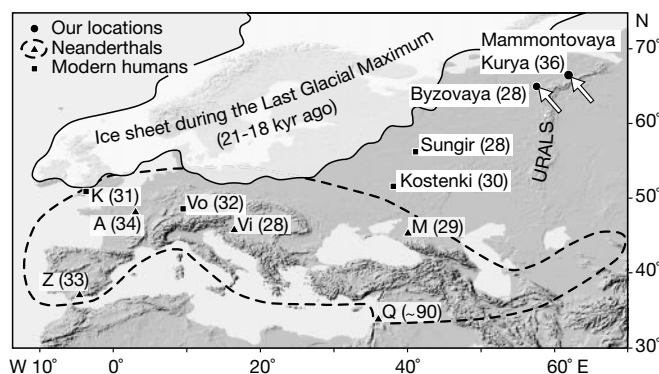
The Mamontovaya Kurya site is located on the southern bank of the Usa river at the Arctic circle (66° 34' N; 62° 25' E), close to the polar Urals (Fig. 1). The riverbed at this site has been known as a place for finding mammoth tusks and bones since the end of the 18th century, but finds of artefacts have not been reported. In order to clarify the stratigraphic context of these bones and to find out if they could be related to early human activities, archaeological and geological field investigations were carried out during the summer seasons of 1992, 1994, 1996 and 1997.

A rich faunal assemblage and several stone artefacts were uncovered for the basal layers of a 12–13 m high river bluff which is cut into the terrace along a bend in the river (Fig. 2). The finds, which were scattered throughout the excavated area (48 m<sup>2</sup>) without any clear concentrations, were incorporated in cross-bedded gravel and sand that accumulated on the floor of an old river channel. Many of the bones uncovered were encapsulated in silt and we also noticed frequent mud clasts within the basal part of the find-bearing channel deposit, which probably reflects slumping from an ancient river terrace covered by over-bank mud. In all, 123 mammalian bones, primarily mammoth (114), but also horse (2), reindeer (5) and wolf (2), were collected (Table 1). The most important find was a 1.3-m-long tusk from a young, 6–8-year-old female mammoth which exhibits a series of distinct grooves (Figs 3 and 4). The marks are 1–2 mm deep, 0.5–1 cm long and appear as densely spaced rows of lines lying crosswise along the tusk. Microscopic analysis reveals that the grooves were made by chopping with a sharp stone edge, unequivocally the work of humans. It is uncertain whether the marks were formed during processing while using the tusk as an anvil, or if they reflect intentional marks with artistic or symbolic meaning. The stone artefacts that were excavated from the same strata comprise five unmodified stone flakes, a straight side-scraper

on a massive cortical blade and a bifacial tool (Fig. 3). The edges of the stone artefacts are sharp and the tusks and bones show minimal signs of wear, indicating a very short transportation and that the material were swiftly buried by alluvial deposits. The few artefacts are not diagnostic and resemble Middle Palaeolithic Mousterian as well as the earliest Upper Palaeolithic assemblages in eastern Europe<sup>5</sup>, a time interval which is also in accordance with the radiocarbon dates discussed below. Similar bifaces are reported for Late Mousterian sites on Crimea, for instance Zaskalnaya V (ref. 6), but they are also known from early Upper Palaeolithic complexes in Eastern Europe, among them Kostenki XII at the Don river<sup>7</sup>. However, we are not able to determine the cultural affiliation on the basis of the sparse material found.

The bones and tusks were in good condition, well suited for radiocarbon dating. The tusk with incision marks was radiocarbon dated to ~36,660 <sup>14</sup>C years before present (yr BP) and three other bones from the same unit yielded similar ages in the range of 34,400–37,400 yr BP (Table 2). This time interval is close to the maximum limit for obtaining accurate radiocarbon dates and the calculated standard deviations for age determinations using conventional dating techniques are normally larger than for accelerator mass spectrometry (AMS) dates. Considering that relatively large amounts of contamination by 'old' inactive carbon is needed to significantly affect the radiocarbon dates, it seems unlikely that the animal remains are significantly younger than the obtained ages. All five radiocarbon dates of various animal remains from the same strata indicate very similar ages. We think it very likely that the artefacts from this layer are of the same age as the tusk and the bones, because the find-bearing strata were buried by several metres of sediment soon after their deposition. Terrestrial plant remains from a slumped mud clast within the find-bearing sand and gravel were dated to ~31,380 and ~30,160 yr BP by using an AMS technique, indicating that the alluvial formation is younger than the bones.

The find-bearing strata is covered by thick layers of cross-bedded sand followed by ripple- and planar-laminated mud, which together are interpreted as a point-bar sequence (arcuate ridge deposit) that accumulated along the inner bend of a meandering river by the addition of individual accretion accompanying migration of the channel. Then follows a 6–10-m-thick formation of diffusely laminated aeolian (wind-driven) silt and sand, in contrast to the pronounced stratified strata below. A series of eight AMS dates of terrestrial plant remains from the alluvial sediments covering the



**Figure 1** Map showing the location of the Palaeolithic sites Mamontovaya Kurya and Byzovaya discussed in the text and the maximum extent of the Eurasian ice sheets during the Last Glacial Maximum (21,000–18,000 yr BP)<sup>10</sup>. The area within which Neanderthal remains have been found is indicated with a dotted line<sup>25</sup>. The location of radiocarbon-dated European sites with skeletal remains of late Neanderthals and early modern humans are also shown<sup>20</sup>: A, Arcy-sur-Cure; K, Kent's Cavern; M, Mazmaiskaya; Vi, Vindija; Vo, Vogelherd; Z, Zafarraya; and Q, Qufzeq<sup>27</sup>. Numbers in parentheses indicate radiocarbon ages ( $\times 10^3$  yr BP).

find-bearing strata yielded ages ranging between  $\sim 31,420$  and  $\sim 23,860$  yr BP whereas optical stimulated luminescence (OSL) dates from the aeolian sediments above give consistently younger ages ranging from  $\sim 19,900$  to  $\sim 13,800$  calendar years BP.

The sedimentological and stratigraphic evidence suggests the following geological history for Mamontovaya Kurya: (1) The refuse of the human occupation was left on the flood plain at around 36,000 yr BP and was shortly thereafter covered by sediments. (2) Slightly before 27,000 yr BP the meandering river undercut these strata and bones and artefacts slumped into the river where they were concentrated in the channel gravel. (3) The bone-bearing gravel was quickly buried by alluvial point-bar deposits as the meander-loop migrated across the site. (4) Aeolian loess-like sediments accumulated on top of the alluvial deposits during the final stage of the Ice Age from  $\sim 20,000$  to  $\sim 13,000$  calendar years ago. (5) Finally the Usa river incised into the terrace during the Holocene and exposed the bones and artefact-bearing layer.

The bone material from Mamontovaya Kurya indicates that humans preyed on, or at least utilized, large herbivorous animals, mostly mammoths. Pollen analysis of the alluvial silt clasts that were found in association with the bones reflects a treeless steppe environment dominated by herbs and grasses, presumably with local stands of willow scrubs (*Salix* spp) along the river banks<sup>8</sup>. Human occupation probably occurred during a relatively mild interlude of the last Ice Age, although the climate at this time was probably considerably colder and more continental than today. This mild interlude may correspond with the Hengelo interstadial (39,000–36,000 yr BP) in western Europe<sup>9</sup>. A palaeo-environmental reconstruction<sup>9</sup> suggests that the landscapes in The Netherlands and northern Germany and eastwards were then covered by a shrub tundra. The northern rim of the Eurasian continent was evidently not glaciated<sup>10</sup> and probably only small mountain glaciers existed in

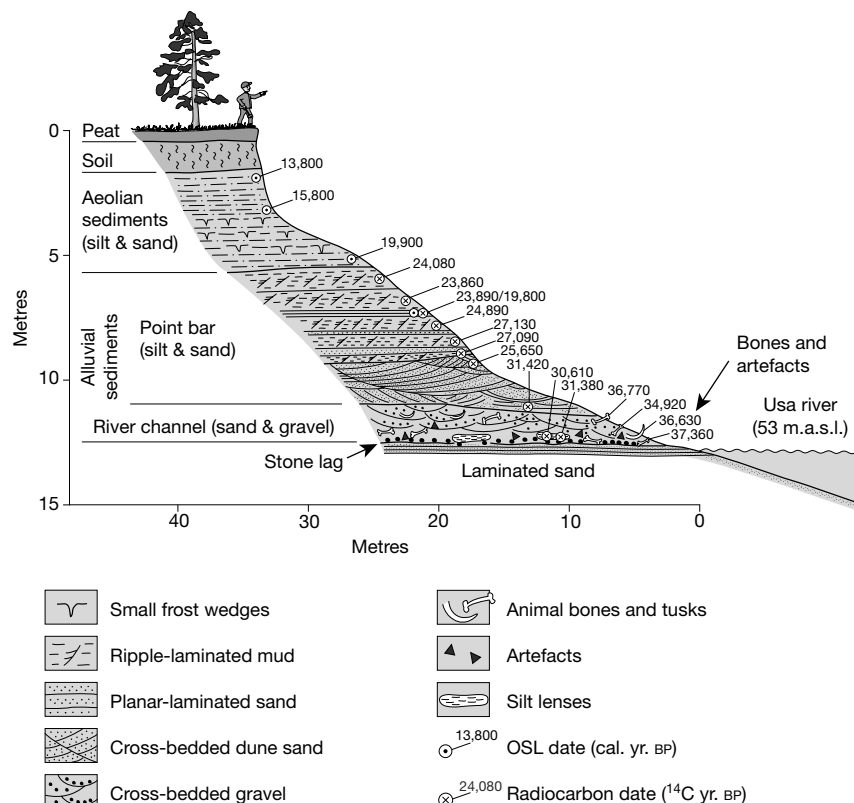
**Table 1 List of bones from Mamontovaya Kurya**

| <i>Mammuthus primigenius</i> Blum<br>(woolly mammoth) | <i>Rangifer tarandus</i> L.<br>(reindeer) | <i>Canis lupus</i> L.<br>(wolf) | <i>Equus caballus</i><br>(horse) |
|-------------------------------------------------------|-------------------------------------------|---------------------------------|----------------------------------|
| 7 ribs                                                | 1 antler                                  | 1 metacarpal                    | 2 teeth                          |
| 1 pelvis                                              | 1 pelvis                                  | 1 unspecified                   |                                  |
| 2 tusks                                               | 1 shoulder                                |                                 |                                  |
| 1 lower jaw                                           | 2 unspecified                             |                                 |                                  |
| 1 skull fragment                                      |                                           |                                 |                                  |
| 3 teeth, upper jaw                                    |                                           |                                 |                                  |
| 2 vertebrae                                           |                                           |                                 |                                  |
| 70 unspecified mammoth                                |                                           |                                 |                                  |

The table shows animal remains collected from the excavated site that could be identified to species. An additional 27 bone fragments could not be identified, but most of them are probably of mammoth.

the Ural Mountains<sup>11,12</sup>. The Scandinavian ice sheet was probably much smaller than during the Last Glacial Maximum some 20,000 yr BP (Fig. 1).

The fact that humans were present in this area as early as around 36,000 yr BP leads us to reassess the history of the earliest human occupation in the Arctic. Until now, the oldest known Palaeolithic sites in the Eurasian Arctic are dated to 13,000–14,000 yr BP<sup>3,4,13</sup>. However, there is an early Upper Palaeolithic site close to the Byzovaya village along the Pechora river, approximately 300 km to the southwest of Mamontovaya Kurya (Fig. 1). At this site nearly 300 artefacts and more than 4,000 animal bones (mainly of mammoth) have been unearthed during several excavations<sup>12,14–17</sup>. The lithic industry of Byzovaya is classified as eastern Szeletien with Aurignacian traits<sup>15,17</sup>, which is typical for many sites of the early Upper Palaeolithic in Eastern Europe<sup>5,18</sup>. An early Upper Palaeolithic age has recently been supported by 13 radiocarbon dates on bones from the find-bearing layer which have yielded ages in the range of 26,000–29,000 yr BP with a mean of  $\sim 28,000$  yr BP<sup>12</sup>.



**Figure 2** The excavated sediment section at Mamontovaya Kurya on the southern bank of the Usa river. The artefacts and bones were uncovered from the river channel deposits near the base of the exposure. Radiocarbon and optically stimulated

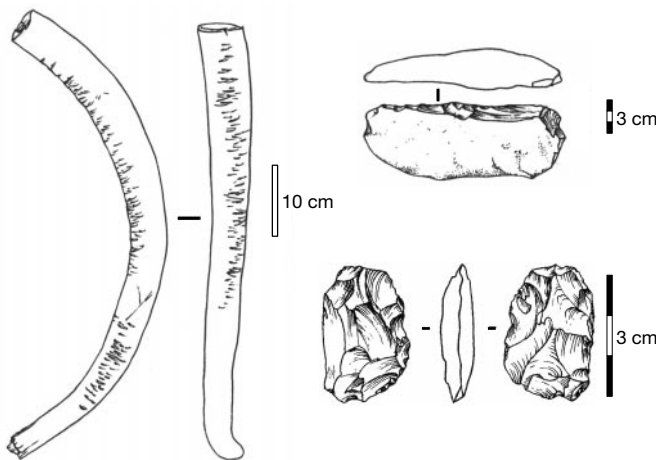
luminescence (OSL) dates from the various layers are indicated (Table 2). We note that the radiocarbon dates are given in  $^{14}\text{C}$  years BP, whereas the OSL dates are in principal calendar years before the present. m.a.s.l., metres above sea level.

We believe that survival of humans in this arctic environment on a year-round basis would have required long-term planning and an extended social network, qualities that are generally associated with modern human behaviour<sup>1</sup>. A pressing question is whether the pioneers who lived in these northern landscapes were members of the ancient Neanderthal population (*Homo sapiens neanderthalensis*) or newcomers from the south. Most scholars associate the Aurignacian industry—the more advanced stone-tool technology that appeared in Europe at around 40,000 yr BP—with the emergence of modern humans<sup>19</sup>. However, the earliest indisputable remains of humans with a fully modern morphology (*Homo sapiens sapiens*) date to 30,000–35,000 yr BP<sup>20</sup>; that is, well after the archaeologically defined transition from the Middle to the Upper Palaeolithic. In European Russia, well preserved skeletons from the famous Palaeolithic site of Sungir, northeast of Moscow (Fig. 1), show that anatomically modern humans were present there not later than ~28,000 yr BP<sup>21,22</sup>. At the Kostenki IV site on the west bank of the Don river, bones of modern humans have been uncovered from strata dated to ~30,000 yr BP<sup>22</sup>. The stone-working technology reflected in the Byzovaya material is similar to that of Sungir and other early Upper Palaeolithic sites of the eastern Szeletien tradition,

indicating that these artefacts were manufactured by modern humans. However, whether the person who inflicted the marks on the tusk from Mamontovaya Kurya, as much as 8,000–9,000 years earlier, belonged to the same human lineage as the residents at Byzovaya and other Palaeolithic sites further to the south is more



**Figure 4** Photograph of the mammoth tusk from Mamontovaya Kurya. The marks appear to have been inflicted by a sharp stone tool.



**Figure 3** Drawings of the mammoth tusk with human-made marks, a side-scraper and a bifacial stone tool (knife?) that were uncovered from the excavated river channel deposits at the Mamontovaya Kurya section.

**Table 2** Optically stimulated luminescence and radiocarbon dates

| Depth below terrace (m) | Laboratory number | Age                        | Dating method   | $\delta^{13}\text{C}$ (‰) | Material dated                   |
|-------------------------|-------------------|----------------------------|-----------------|---------------------------|----------------------------------|
| 2.0                     | 99253-1           | 13,800 $\pm$ 1,100†        | OSL             | –                         | Aeolian sand                     |
| –                       | 99253-2           | 14,400 $\pm$ 900†          | OSL             | –                         | Aeolian sand (adjacent section)  |
| 3.3                     | 99253-3           | 15,800 $\pm$ 1,000†        | OSL             | –                         | Aeolian silt                     |
| 4.6                     | 99253-4           | 19,900 $\pm$ 1,300†        | OSL             | –                         | Aeolian silt                     |
| 7.4                     | 99253-0           | 19,800 $\pm$ 2,100†        | OSL             | –                         | Alluvial silt                    |
| 6.0                     | ETH-20830*        | 24,080 $\pm$ 220‡          | <sup>14</sup> C | –25.9                     | Terrestrial moss                 |
| 6.9                     | Beta-11950*       | 23,860 $\pm$ 120‡          | <sup>14</sup> C | –25.8                     | Terrestrial moss                 |
| 7.4                     | Beta-119502*      | 23,890 $\pm$ 140‡          | <sup>14</sup> C | –26.0                     | Terrestrial moss                 |
| 7.9                     | ETH-20831*        | 24,890 $\pm$ 210‡          | <sup>14</sup> C | –25.1                     | Terrestrial moss                 |
| 8.5                     | Beta-4072*        | 27,130 $\pm$ 180‡          | <sup>14</sup> C | –26.4                     | Terrestrial moss                 |
| 8.9                     | ETH-20832*        | 27,090 $\pm$ 240‡          | <sup>14</sup> C | –24.4                     | Terrestrial moss                 |
| 9.4                     | TUa-1514*         | 25,650 $\pm$ 535‡          | <sup>14</sup> C | –28.2                     | Terrestrial plants               |
| 10.9                    | ETH-21437*        | 31,420 $\pm$ 370‡          | <sup>14</sup> C | –21.0                     | Terrestrial moss                 |
| 11.7                    | T-11503           | 36,770 $\pm$ 2,620/–1,980‡ | <sup>14</sup> C | –                         | Horse tooth                      |
| 12.0                    | ETH-21439*        | 30,610 $\pm$ 350‡          | <sup>14</sup> C | –18.9                     | Terrestrial moss                 |
| 12.1                    | ETH-21438*        | 31,380 $\pm$ 380‡          | <sup>14</sup> C | –22.2                     | Terrestrial moss                 |
| 12.1                    | T-11403           | 36,630 $\pm$ 1,310/–1,130‡ | <sup>14</sup> C | –                         | Mammoth tusk with marks          |
| 12.1                    | T-11504           | 34,360 $\pm$ 630‡          | <sup>14</sup> C | –                         | Mammoth bone                     |
| 12.1                    | LU-4001           | 37,360 $\pm$ 970‡          | <sup>14</sup> C | –                         | Mammoth bone                     |
| –                       | LU-3994           | 34,920 $\pm$ 1,040‡        | <sup>14</sup> C | –                         | Mammoth tusk (uncertain context) |

\* AMS date.

† Calendar yr BP.

‡ <sup>14</sup>C yr BP.

The OSL dates (calendar years), measured on quartz grains in the sand grain fraction, were produced at the Nordic Laboratory for Luminescence Dating, Risø National Laboratory, Denmark. The radiocarbon dates (<sup>14</sup>C yr BP) were carried out at various laboratories. Beta, Beta analytic; ETH, the Swiss Federal Institute of Technology AMS Facility; T, Trondheim Radiocarbon Laboratory; TUa, prepared at the Trondheim and measured at the accelerator at the Svedberg Laboratory, Uppsala; LU, St Petersburg University.



uncertain. If this person was a modern human who descended from temperate areas, as predicted by the 'Out of Africa' hypothesis<sup>2</sup>, then the Russian Arctic was occupied by *Homo sapiens sapiens* shortly after the first newcomers entered Europe<sup>23,24</sup>. On the other hand, if the person was a Neanderthal, then these humans expanded much further north than hitherto assumed, implying that their stage of cultural development was not a barrier to colonization of this Arctic habitat. Whoever she or he was, the findings from Mamontovaya Kurya provide evidence that the European part of the Arctic was inhabited by humans long before the Neanderthals vanished from the continent soon after 28,000 yr BP<sup>20,25,26</sup>. □

Received 27 February; accepted 27 June 2001.

- Gamble, C. *Paleolithic Societies of Europe* 268–426 (Cambridge Univ. Press, Cambridge, 1999).
- Stringer, C. B. & Mackie, R. *African Exodus: the Origin of Modern Humanity* 84–111 (Cape, London, 1996).
- Hoffecker, J. F., Powers, W. R. & Goebel, T. The colonization of the Beringia and the peopling of the New World. *Science* **259**, 46–53 (1993).
- Mochanov, Yu. A. *Initial Settling of the Territory of North-Eastern Asia* (Nauka, Novosibirsk, 1977) (in Russian).
- Allsworth-Jones, P. *The Szeletian* 83–198 (Clarendon, Oxford, 1986).
- Kolosov, Yu. G. *Mousterian Sites of the Belogorsk Area, Crimea* (Nauka, Kiev, 1983) (in Russian).
- Praslov, N. D. & Rogachev, A. N. (eds) *Paleolit Kostenkovsko–Borschevskogo raiona na Donu 1879–1979 (Palaeolithic of the Kostenki–Borshevo Area on the Don River)* 16–66 (Nauka, Leningrad, 1982) (in Russian).
- Halvorsen, L. S. *Palaeovegetation and Environment during Weichselian Stadials and Interstadials at Mamontovaya Kurja and Sokolova in the Pechora Basin, Northern Russia*. Thesis, Univ. Bergen (2000).
- Van Andel, T. H. & Tzedakis, P. C. Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: An overview. *Quat. Sci. Rev.* **15**, 481–500 (1996).
- Svendsen, J. I. et al. Maximum extent of the Eurasian ice sheet in the Barents and Kara Sea region during the Weichselian. *Boreas* **28**, 234–242 (1999).
- Astakhov, V. I. et al. Marginal formations of the last Kara and Barents ice shelves in northern European Russia. *Boreas* **28**, 23–45 (1999).
- Mangerud, J., Svendsen, J. I. & Astakhov, V. I. Age and extent of the Barents and Kara ice sheets in Northern Russia. *Boreas* **28**, 46–80 (1999).
- Powers, W. R. in *Humans at the End of the Ice Age—The Archaeology of the Pleistocene–Holocene Transition* (eds Straus, L. G., Eriksen, B. V., Erlandson, J. M. & Yesner, D. R.) 229–253 (Plenum, New York, 1996).
- Kanivets, V. I. *The Paleolithic of the Extreme North-East of Europe* (Nauka, Moscow, 1976) (in Russian).
- Pavlov, P. Yu. The excavation of Byzovaya Upper Palaeolithic site in 1983–1985 in Pamiatniki materialnoj kultury na Evropeiskom severo-vostoke 7–16 (Russian Acad. Sci., Syktyvkar, 1986) (in Russian).
- Pavlov, P. Yu. *The Palaeolithic Archaeology of the Komi Republic* 44–91 (DiK, Moskva, 1996) (in Russian).
- Pavlov, P. Yu. & Indrelid, S. Initial settling of North-Eastern Europe. *Bull. Inst. Biol., Syktyvkar* **2**, 24–26 (1999).
- Anikovich, M. V. *Early Upper Palaeolithic in Eastern Europe* (AN SSSR, St Petersburg, 1991) (in Russian).
- Straus, L. G. Age of Modern Europeans. *Nature* **342**, 476–477 (1989).
- Smith, F. H., Trinkhaus, E., Pettitt, P. B., Karavanic, I. & Paunovic, M. Direct radiocarbon dates for Vindija G1 and Velika Pecina Late Pleistocene hominid remains. *Proc. Natl Acad. Sci. USA* **96**, 12281–12286 (1999).
- Bader, O. N. *Sungir—Upper Palaeolithic Site* (Nauka, Moskva, 1978) (in Russian).
- Sinitsyn, A. A. & Praslov, N. D. *Radiocarbon chronology of the Paleolithic of Eastern Europe and Northern Asia. Problems Perspectives* (Russian Acad. Sci., St Petersburg, 21–66 1997).
- Bocquet-Appel, J.-P. & Demars, P. Y. Neanderthal contraction and northern human colonization of Europe. *Antiquity* **74**, 544–552 (2000).
- Straus, L. G., Bicho, N. & Winegardner, A. C. The Upper Palaeolithic settlement of Iberia: first-generation maps. *Antiquity* **74**, 553–566 (2000).
- Ovchinnikov, I. V. et al. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**, 490–493 (2000).
- d'Erico, F., Zilhao, J., Julien, M., Baffier, D. & Pelegrin, J. Neanderthal acculturation in Western Europe? A critical review of evidence and its interpretation. *Curr. Anthropol.* **39** (Suppl.), 1–44 (1998).
- Bar-Yosef, O. in *Neanderthals and Modern Humans in Western Asia* (eds Akazawa, T., Aoki, K. & Bar-Yosef, O.) 39–56 (Plenum, New York, 1998).

## Acknowledgements

This work is a contribution to the Russian–Norwegian research project Paleo Environment and Climate History of the Russian Arctic (PECHORA), which forms part of the European Science Foundation Program Quaternary Environment of the Eurasian North (QUEEN). We thank E. Giria, Institute of History of the Material Culture, St Petersburg University, for carrying out microscopic analysis of the marks on the mammoth tusk. We thank J. Mangerud for his comments on this manuscript and for discussions. The bond material was identified by I. Kuzmina and D. Ponomarev. The drawing of the tusk and the stone artefacts were done by N. Pavlov and the figures by E. Bjørseth. We thank the Norwegian Research Council for financial support.

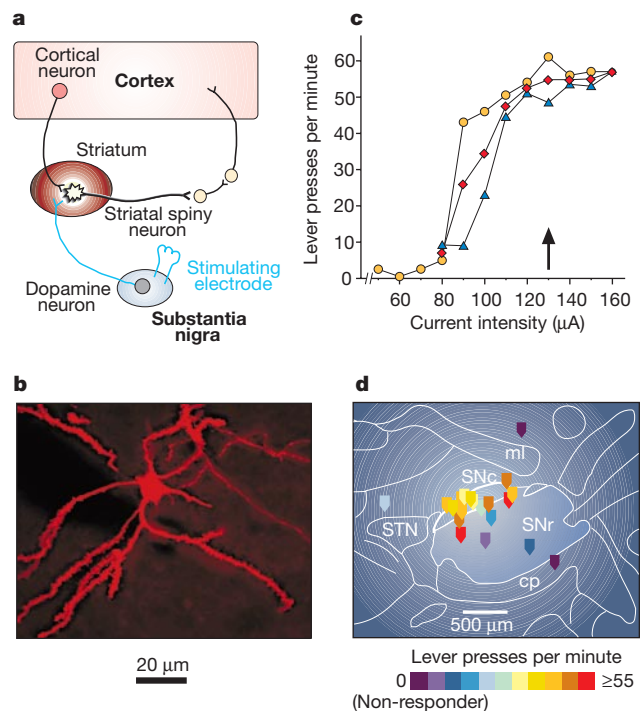
Correspondence and requests for materials should be addressed to J.I.S. (e-mail: john.svendsen@smr.uib.no).

# A cellular mechanism of reward-related learning

John N. J. Reynolds\*†‡, Brian I. Hyland\*‡ & Jeffery R. Wickens\*†

\* The Neuroscience Research Centre; and † Department of Anatomy and Structural Biology; and ‡ Department of Physiology, University of Otago, School of Medical Sciences, Dunedin, New Zealand

Positive reinforcement helps to control the acquisition of learned behaviours. Here we report a cellular mechanism in the brain that may underlie the behavioural effects of positive reinforcement. We used intracranial self-stimulation (ICSS) as a model of reinforcement learning<sup>1</sup>, in which each rat learns to press a lever that applies reinforcing electrical stimulation to its own substantia nigra<sup>2,3</sup>. The outputs from neurons of the substantia nigra terminate on neurons in the striatum in close proximity to inputs from the cerebral cortex on the same striatal neurons<sup>4</sup>. We measured the effect of substantia nigra stimulation on these inputs from the cortex to striatal neurons and also on how quickly the rats learned to press the lever. We found that stimulation of the substantia nigra (with the optimal parameters for lever-pressing behaviour) induced potentiation of synapses between the cortex and the striatum, which required activation of dopamine receptors. The degree of potentiation within ten minutes of the ICSS trains was correlated with the time taken by the rats to learn ICSS behaviour.



**Figure 1** Intracranial self-stimulation of the nigrostriatal system. **a**, Overview of the circuit studied. **b**, Confocal micrograph of a striatal spiny neuron injected with biocytin during intracellular recording (streptavidin-Texas Red label). **c**, Lever-pressing rate for one rat in response to increments (yellow circles) and decrements (blue triangles) in substantia nigra stimulus intensity. Arrow indicates the optimal current that just maximized the average rate (red diamonds). **d**, Approximate midpoint of the final stimulating electrode positions (sagittal section at mediolateral +1.9 mm; ref. 30). Arrowheads coded by maximum lever-pressing rate. SNr, substantia nigra pars reticulata; STN, subthalamic nucleus; cp, cerebral peduncle; ml, medial lemniscus.

We propose that stimulation of the substantia nigra when the lever is pressed induces a similar potentiation of cortical inputs to the striatum, positively reinforcing the learning of the behaviour by the rats.

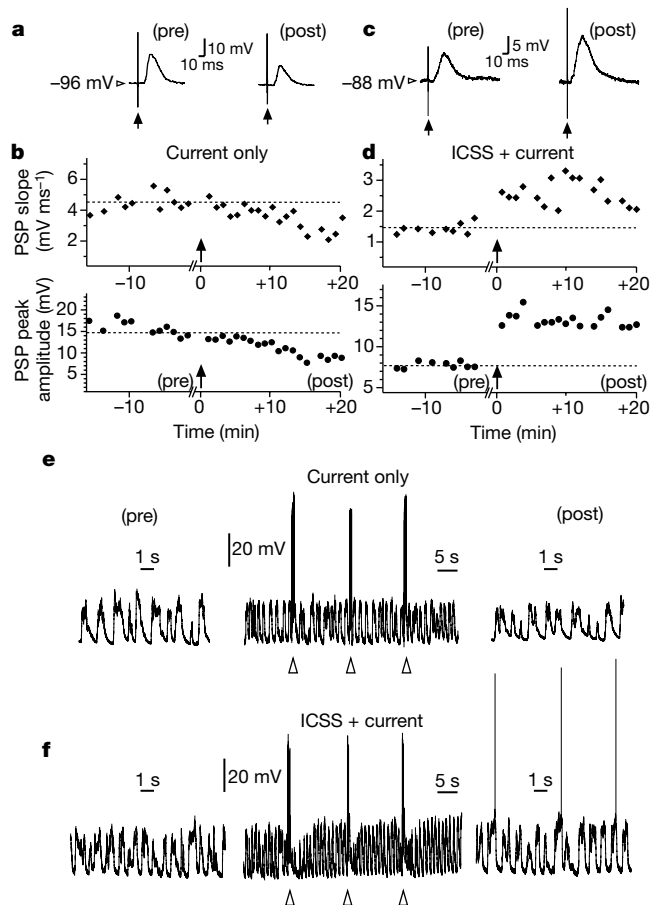
Recent electrophysiological studies have identified circuits in the brain which carry specific signals about past and future rewards<sup>5</sup>. The dopamine neurons in the pars compacta of the substantia nigra show short, phasic activation after the presentation of food or liquid rewards<sup>6,7</sup> and are believed to be important in reward-related learning. Nigral dopamine neurons project predominantly to the striatum, where changes in neural activity associated with reward-related learning have been described<sup>8,9</sup>. Determining the effect of short, phasic activation of the dopamine neurons on neural information processing in the striatum is, therefore, a crucial step towards understanding the cellular mechanism of reward-related learning.

We first established ICSS behaviour in male rats using stimulating electrodes permanently implanted in the substantia nigra (Fig. 1a, d). In these behavioural experiments we measured the time taken to establish ICSS (range, 2 to 240 min) and determined the optimal current intensity<sup>10,11</sup> (range, 90 to 275  $\mu$ A; Fig. 1c) and maximum lever-pressing rate (range, 18 to 56 presses per min) for each animal. Over the following two days, we extinguished the newly learned

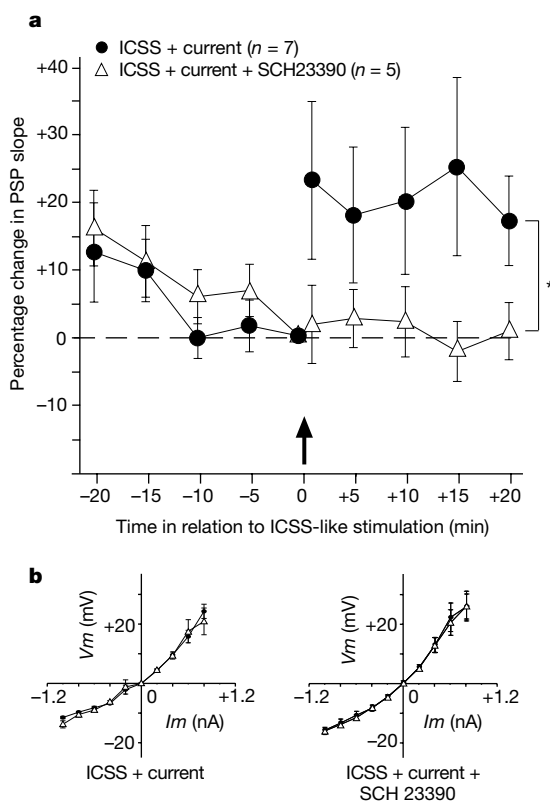
behaviour by dissociating lever-pressing from substantia nigra stimulation.

After the behavioural sessions, the same animals were anaesthetized and prepared for *in vivo* electrophysiological recording. Intracellular records were obtained from striatal neurons identified as spiny projection neurons on the basis of typical membrane potential fluctuations<sup>12,13</sup> (Fig. 2e, f), verified in some cases by histological examination (Fig. 1b). The efficacy of corticostriatal synaptic inputs to these neurons was measured from post-synaptic potentials (PSPs; Fig. 2a, c) evoked by control stimulation of the contralateral cerebral cortex<sup>14,15</sup>. These measurements were made at regular intervals before and after ICSS-like stimulation.

In the electrophysiological experiments, ICSS-like stimulation was applied to the substantia nigra electrodes. Stimulus trains contained the same number of pulses and were delivered at the same frequency as used for ICSS, and for each recorded cell were of the current intensity found to be most effective at supporting ICSS behaviour for each individual rat. Activity of the corticostriatal pathway during ICSS-like stimulation was provided by spontaneous cortical input. This input causes large-amplitude membrane potential fluctuations in striatal spiny cells<sup>12</sup>; however, spiny neurons are highly polarized and a proportion of them do not fire action potentials spontaneously<sup>16</sup>. Therefore, to simulate the phasic activation of striatal neurons that occurs during movement<sup>17,18</sup>, we paired the substantia nigra stimulation with an intracellular current injection sufficient to induce action potential firing. To control for the effect of current injection and spontaneous corticostriatal input, a comparable level of intracellular current



**Figure 2** Effect of treatment protocols on corticostriatal responses. Data is shown from representative striatal neurons in a control rat (**a, b, e**) and a rat experienced in intracranial self-stimulation (ICSS) (**c, d, f**). **a, c**, Post-synaptic potentials (1-min average) just before and 20 min after intracellular current injection alone (**a**) or in conjunction with ICSS-like stimulation (**c**). Arrows indicate stimulus artefact. **b, d**, Slope (diamonds) and peak amplitude (circles) of the post-synaptic potential (PSP) (1-min averages) in relation to application of the protocols (arrows). **e, f**, Membrane potential fluctuations of the neurons before, during and 20 min after the protocols. Protocol application indicated by open arrowheads (action potentials truncated).



**Figure 3** Group average effect of ICSS-like stimulation on corticostriatal responses and cellular properties. **a**, Changes in PSP slope (as percentage of baseline) in both groups that received ICSS-like stimulation. Baseline (dashed line) is average of 5 min of PSPs before application of stimulus protocol (arrow). Asterisk, significantly different between ICSS (no drug) and the ICSS (SCH 23390) group at +20 min ( $P < 0.02$ ). **b**, Group average current-voltage relation before (circles) and 20 min after (triangles) the protocols. The voltage at each point is normalized to membrane potential at zero current. All graph points are group mean  $\pm$  standard error of the mean, s.e.m.

**Table 1 Cellular properties of recorded neurons**

|                                           | Time* | ICSS+<br>current<br>(no drug)† | ICSS+<br>current+<br>SCH 23390‡ | Current<br>only† |
|-------------------------------------------|-------|--------------------------------|---------------------------------|------------------|
| Input resistance (MΩ)                     | Pre   | 26.0 ± 6.8                     | 30.5 ± 11.9                     | 25.0 ± 8.9       |
|                                           | Post  | 24.2 ± 8.8                     | 31.3 ± 12.4                     | 24.0 ± 9.0       |
| Membrane potential (mV)                   | Pre   | -90.1 ± 3.3                    | -91.1 ± 2.1                     | -87.4 ± 8.2      |
|                                           | Post  | -90.9 ± 1.8                    | -91.6 ± 1.6                     | -87.2 ± 7.9      |
| Baseline PSP slope (mV ms <sup>-1</sup> ) |       | 2.46 ± 0.76                    | 3.07 ± 0.50                     | 2.96 ± 1.05      |
| Baseline PSP peak amplitude (mV)          |       | 13.4 ± 5.6                     | 17.5 ± 6.5                      | 15.4 ± 4.4       |
| Protocol depolarization (mV)              |       | -51.3 ± 7.9                    | -57.7 ± 6.8                     | -50.7 ± 6.4      |
| ICSS-like stimulus intensity (μA)‡        |       | 181 ± 66                       | 144 ± 43                        |                  |

Data are mean ± s.d.

\*There were no significant differences between values before 'pre' and 20 min after 'post' application of the treatment protocols (paired *t*-test, *P* > 0.05). All comparisons based on at least five cells per group.

†There were no significant differences between treatment groups (one way ANOVA, *P* > 0.05).

‡The same as the optimal current intensity applied to the substantia nigra during ICSS.

injection was applied to cells in a group that received no ICSS-like stimulation.

Our results revealed a clear potentiation of corticostriatal synaptic efficacy following stimulation of the substantia nigra with the optimal ICSS parameters. As illustrated in Fig. 2a and b, control cells (current injection and spontaneous corticostriatal input only) showed depression of PSPs. In the control group as a whole, the average effect was depression ( $-15.8 \pm 5.1\%$  at 20 min, *n* = 7). In contrast, ICSS-like stimulation reversed the inherent tendency towards depression, inducing potentiation of up to 97% (Fig. 2c and d). The average effect of ICSS-like stimulation (Fig. 3a) was potentiation of the PSP slope ( $+17.2 \pm 6.7\%$  at 20 min, *n* = 7), which differed significantly from the control group at all time points following the treatment protocols (*P* < 0.001). Changes in synaptic responses in each group were not secondary to changes in general cellular properties, because the latter were not affected by the treatment protocols (Table 1). We conclude that the changes in synaptic responses reflect modulation of synaptic efficacy in the corticostriatal pathway. Although synaptic plasticity has been previously demonstrated in the corticostriatal pathway *in vivo*<sup>15,27</sup> the induction of synaptic potentiation by stimulation that affects behaviour is, to our knowledge, a new finding.

To determine whether the effects of substantia nigra stimulation were mediated by dopamine, the dopamine receptor antagonist SCH 23390 was applied to a group that received ICSS-like stimulation. Potentiation was not observed in these neurons. The average effect of ICSS-like stimulation in the dopamine antagonist group ( $+1.1 \pm 4.3\%$  at 20 min, *n* = 5) was significantly different (*P* < 0.02) from the no-drug ICSS group (Fig. 3a). Cellular properties were not affected by the drug or by ICSS-like stimulation (Fig. 3b and Table 1). These data demonstrate that stimulation of the substantia nigra at a current intensity that is optimal to support

ICSS in the same animal, and using the same electrodes, reliably induces dopamine-dependent potentiation of corticostriatal PSPs. It is probable that the same potentiation would have been induced at corticostriatal synapses during ICSS in an awake animal.

Although the most effective ICSS sites are close to dopamine cell bodies<sup>19</sup> (also see Fig. 1d) the substrate directly activated by the ICSS electrode remains unclear<sup>20,21</sup>. It has been suggested that activation of descending myelinated fibres afferent to the dopamine neurons is required<sup>22</sup>. But our results do indicate that the final outcome of substantia nigra ICSS is dopamine-dependent potentiation. This effect is on corticostriatal synaptic efficacy rather than cortical excitability, because the contralateral cortex used to evoke test responses does not receive a crossed dopaminergic input from the substantia nigra<sup>23</sup>.

Recent theories of reinforcement learning<sup>24,25</sup> propose that dopamine acts heterosynaptically to facilitate synaptic plasticity at corticostriatal synapses. Such potentiation of the corticostriatal pathway can lead to increased probability of firing in striatal output neurons, as shown in Fig. 2f. The corticostriatal synapses are a key control point in a re-entrant loop involving the motor cortex and striatum<sup>26</sup> (Fig. 1a), so dopamine-dependent potentiation of these synapses may serve as a cellular mechanism for the learning of behavioural responses.

Because the intracellular recording and ICSS experiments were performed in the same animals, we were able to investigate whether there was any correlation between the magnitude of changes in synaptic efficacy induced by ICSS-like stimulation and behavioural variables. Our combined behavioural–electrophysiological analysis revealed that the degree of potentiation induced by ICSS-like stimulation was negatively correlated with the time taken to learn ICSS. This was observed at time points tested up to 10 min following the stimulus trains (+1 min: *r* = -0.91, *P* < 0.01; +5 min: *r* = -0.78, *P* < 0.05; +10 min: *r* = -0.80, *P* < 0.05) (Fig. 4a). No correlation was observed at later time points (+15 min: *r* = -0.62; +20 min: *r* = -0.46; not significant, NS), even though potentiation persisted for at least 20 min. In contrast to the correlation with the time taken to learn ICSS, there was no correlation between the degree of potentiation and the maximum lever-pressing rate at any time point (+1 min: *r* = -0.16; +5 min: *r* = -0.28; +10 min: *r* = -0.08; +15 min: *r* = -0.12; +20 min: *r* = 0.15, NS) (Fig. 4b). Thus there is a temporally specific relationship between the changes in corticostriatal synaptic efficacy induced by substantia nigra stimulation and the learning of ICSS behaviour, but not the subsequent performance of the behaviour.

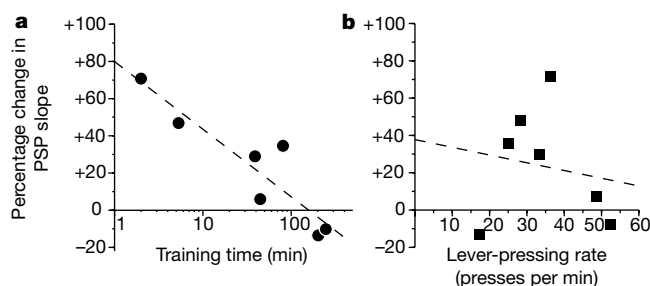
The correlation observed between PSP changes and the rate of acquisition of the ICSS behaviour suggests that enhanced information transmission through the corticostriatal circuit is an important factor in the acquisition of lever-pressing behaviour. We propose that this may be due to selective strengthening of particular pathways through the corticostriatal circuit, as previously proposed<sup>28</sup>. The correlation with the initial degree of potentiation implies that potentiation may influence immediate activities related to achieving ICSS, such as approach to the lever following a priming stimulation. These mechanisms may be generalized to other forms of learning facilitated by positive reinforcement. □

## Methods

### Intracranial self-stimulation methods

All procedures were approved by the University of Otago's Committee on Ethics in the care and Use of Laboratory Animals.

Twenty male Wistar rats (230–385 g) were implanted for ICSS. Rats were anaesthetized with pentobarbitone (60 mg kg<sup>-1</sup> intraperitoneally) and placed in a stereotaxic frame in the flat skull position. Prophylactic antibiotic was administered (Streptin, 0.25 ml subcutaneously) and surgery performed using an aseptic technique. A bipolar stimulating electrode with tips separated by 0.7 mm was positioned at the level of the substantia nigra pars compacta (medial tip aimed for anteroposterior -4.8 mm, mediolateral 1.3 mm relative to bregma; dorsoventral 7.7 mm from brain surface). At least three days were allowed for post-operative recovery.



**Figure 4** Relationship between changes in synaptic efficacy and ICSS learning and performance. **a**, Correlation between degree of potentiation induced at +1 min following ICSS-like stimulus trains and ICSS training time (*r* = -0.91, *P* < 0.01). **b**, No correlation between degree of potentiation induced at +1 min following ICSS-like stimulation and maximum lever-pressing rate (*r* = -0.16, *P* = 0.73). For both graphs, each point is from a single animal in the no-drug ICSS group.



Rats were trained to perform ICSS in a behavioural chamber with a lever which, when pressed, immediately triggered delivery of a single stimulus train (100 Hz, 0.5 ms biphasic pulses, 500 ms train duration) to the substantia nigra electrode. Priming stimulus trains were administered by the experimenter when the animal was close to the lever, in order to facilitate approach and chance contact with the lever. All training was performed using the same training techniques for each animal. The training time was the total period of time that the rat was in the chamber with access to the lever, until lever-pressing commenced (at least 20 self-initiated presses in two consecutive minutes). Two animals did not reach the criterion within five days and were not used in further experimental procedures.

In the remaining ICSS responders ( $n = 18$ ), response rate versus current intensity data was collected daily until the currents required to support minimum and maximum lever-pressing rates (optimal current) were stable to within  $\pm 10\%$  for three consecutive sessions. Each animal then underwent two days of extinction where the animals were allowed free access to the lever with the self-stimulation circuit disabled. In addition, non-contingent stimulus trains were applied by the experimenter at 5-min intervals when the animal was away from the lever, in order to dissociate lever-pressing from the delivery of stimulus trains. Successful extinction was defined as a failure to approach the lever within the interval between two successive primes. This was successfully achieved in all animals.

## Electrophysiological recording methods

Electrophysiological data was obtained from 12 of the 18 ICSS-experienced rats and 5 additional control animals. Each rat was anaesthetized with urethane ( $1.4\text{--}1.6\text{ g kg}^{-1}$  intraperitoneally), supplemented with ketamine ( $10\text{ mg kg}^{-1}$  intramuscularly) and xylazine ( $2\text{ mg kg}^{-1}$  intramuscularly). An electrode was implanted in all animals in the medial agranular cortex contralateral to the recording site (anteroposterior  $+10.6$  to  $12.7\text{ mm}$ , mediolateral  $-2.0\text{ mm}$  relative to interaural line; dorsoventral  $1.6$  to  $2.0\text{ mm}$  from brain surface). In control rats, a bipolar stimulating electrode was positioned in the substantia nigra at the same coordinates as the permanently implanted electrode in the ICSS-experienced rats but this electrode was not used for ICSS-like stimulation.

Intracellular records were made from striatal neurons with microelectrodes pulled from  $3.0\text{-mm}$  diameter glass and filled with  $1\text{ M K-acetate}$  containing  $4\%$  biocytin. Post-synaptic potentials evoked by cortical test stimuli ( $0.1\text{ Hz}$ ,  $0.1\text{ ms}$  biphasic pulses,  $300$  to  $990\text{ }\mu\text{A}$ ) were routinely recorded for  $20\text{ min}$  before and  $20\text{ min}$  after the application of a treatment protocol. In ICSS-experienced rats, this consisted of six ICSS-like stimulus trains administered at ten-second intervals at the animal's optimal current intensity (Table 1).

In five experiments, the dopamine D1/D5 receptor antagonist SCH 23390 ( $40\text{ }\mu\text{g kg}^{-1}$  intraperitoneally in  $0.1\text{ ml } 0.9\%$  NaCl) was administered approximately  $30\text{ min}$  before the application of ICSS-like stimulus trains. This dose was based on our preliminary behavioural experiments (data not shown) and the reports of others<sup>29</sup>, which show it is sufficient to attenuate the rewarding effect of ICSS. Each ICSS-like stimulus train was paired with a current pulse just above the threshold for action potential firing ( $1.3 \pm 0.5\text{ nA}$ ) of similar intensity to control rats that received only current injection ( $1.1 \pm 0.2\text{ nA}$ ). Cellular properties of the recorded neurons were measured as previously described<sup>15</sup>. Post-synaptic potentials were compared between treatment groups by a repeated measures analysis of variance (ANOVA) (treatment effect: degrees of freedom,  $d.f. = 2$ ,  $F = 5.1$ ,  $P < 0.05$ ; no time effect or interaction). Post-hoc tests between individual groups were performed using a Bonferroni analysis (overall level of significance  $P < 0.05$ ). Substantia nigra electrode positions were verified in unstained vibratome sections. There were no systematic differences in electrode positions between groups.

Received 29 May; accepted 12 July 2001.

- Olds, J. & Milner, P. Positive reinforcement produced by electrical stimulation of septal and other regions of the rat brain. *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
- Beninger, R. J., Bellisle, F. & Milner, P. M. Schedule control of behavior reinforced by electrical stimulation of the brain. *Science* **196**, 547–549 (1977).
- Major, R. & White, N. Memory facilitation by self-stimulation reinforcement mediated by the nigro-neostriatal bundle. *Physiol. Behav.* **20**, 723–733 (1978).
- Smith, A. D. & Bolam, J. P. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci.* **13**, 259–265 (1990).
- Schultz, W. Multiple reward signals in the brain. *Nature Rev. Neurosci.* **1**, 199–207 (2000).
- Mirenovic, J. & Schultz, W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* **72**, 1024–1027 (1994).
- Mirenovic, J. & Schultz, W. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* **379**, 449–451 (1996).
- Matsumoto, N., Hanakawa, T., Maki, S., Graybiel, A. M. & Kimura, M. Role of nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. *J. Neurophysiol.* **82**, 978–998 (1999).
- Aosaki, T., Graybiel, A. M. & Kimura, M. Effect of the nigrostriatal dopamine system on acquired neural responses in the striatum of behaving monkeys. *Science* **265**, 412–415 (1994).
- Hodos, W. & Valenstein, E. S. An evaluation of response rate as a measure of rewarding intracranial stimulation. *J. Comp. Physiol. Psychol.* **55**, 80–84 (1962).
- Reynolds, R. W. The relationship between stimulation voltage and rate of hypothalamic self-stimulation in the rat. *J. Comp. Physiol. Psychol.* **51**, 193–198 (1958).
- Wilson, C. J. & Kawaguchi, Y. The origins of two-state spontaneous membrane potential fluctuations of neostriatal spiny neurons. *J. Neurosci.* **16**, 2397–2410 (1996).
- Stern, E. A., Jaeger, D. & Wilson, C. J. Membrane potential synchrony of simultaneously recorded striatal spiny neurons *in vivo*. *Nature* **394**, 475–478 (1998).
- Wilson, C. J. Postsynaptic potentials evoked in spiny neostriatal projection neurons by stimulation of ipsilateral and contralateral neocortex. *Brain Res.* **367**, 201–213 (1986).

- Reynolds, J. N. J. & Wickens, J. R. Substantia nigra dopamine regulates synaptic plasticity and membrane potential fluctuations in the rat neostriatum, *in vivo*. *Neuroscience* **99**, 199–203 (2000).
- Wilson, C. J. & Groves, P. M. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* **220**, 67–80 (1981).
- Alexander, G. E. & Crutcher, M. D. Preparation for movement: neural representations of intended direction in three motor areas of the monkey. *J. Neurophysiol.* **64**, 133–150 (1990).
- Kimura, M. Behaviorally contingent property of movement-related activity of the primate putamen. *J. Neurophysiol.* **63**, 1277–1296 (1990).
- Corbett, D. & Wise, R. A. Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable electrode mapping study. *Brain Res.* **185**, 1–15 (1980).
- Fibiger, H. C., LePiane, F. G., Jakubovic, A. & Phillips, A. G. The role of dopamine in intracranial self-stimulation of the ventral tegmental area. *J. Neurosci.* **7**, 3888–3896 (1987).
- Wise, R. A. Addictive drugs and brain stimulation reward. *Annu. Rev. Neurosci.* **19**, 319–340 (1996).
- Bielajew, C. & Shizgal, P. Evidence implicating descending fibers in self-stimulation of the medial forebrain bundle. *J. Neurosci.* **6**, 919–929 (1986).
- Swanson, L. W. The projections of the ventral tegmental area and adjacent regions: a combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain Res. Bull.* **9**, 321–353 (1982).
- Suri, R. E. & Schultz, W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91**, 871–890 (1999).
- Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
- Alexander, G. E., DeLong, M. R. & Strick, P. L. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* **9**, 357–381 (1986).
- Chapier, S. & Deniau, J. M. In vivo activity-dependent plasticity at cortico-striatal connections: evidence for physiological long-term potentiation. *Proc. Natl Acad. Sci. USA* **94**, 7036–7040 (1997).
- Miller, R. M. *Meaning and Purpose in the Intact Brain* 62–120 (Clarendon Press, Oxford, 1981).
- Hunt, G. E., Atrens, D. M. & Jackson, D. M. Reward summation and the effects of dopamine D1 and D2 agonists and antagonists on fixed-interval responding for brain stimulation. *Pharmacol. Biochem. Behav.* **48**, 853–862 (1994).
- Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Academic, New York, 1998).

## Acknowledgements

We thank A. Kean and C. Booker for technical assistance. This work was supported by the Health Research Council of New Zealand, Lottery Health Research and the New Zealand Neurological Foundation.

Correspondence and requests for materials should be addressed to J.R.W. (e-mail: jeff.wickens@stonebow.otago.ac.nz).

# Mobilization of a *Drosophila* transposon in the *Caenorhabditis elegans* germ line

Jean-Louis Bessereau\*, Ashley Wright, Daniel C. Williams, Kim Schuske, M. Wayne Davis & Erik M. Jorgensen

Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, Utah 84112-0840, USA

Transposons have been enormously useful for genetic analysis in both *Drosophila* and bacteria. Mutagenic insertions constitute molecular tags that are used to rapidly clone the mutated gene. Such techniques would be especially advantageous in the nematode *Caenorhabditis elegans*, as the entire sequence of the genome has been determined. Several different types of endogenous transposons are present in *C. elegans*, and these can be mobilized in mutator strains (reviewed in ref. 1). Unfortunately, use of these native transposons for regulated transposition in *C. elegans* is limited. First, all strains contain multiple copies of these transposons and thus new insertions do not provide unique tags. Second, mutator strains tend to activate the transposition of several classes of transposons, so that the type of transposon associated with a particular mutation is not known. Here we demonstrate that the *Drosophila* mariner element *Mos1* can be mobilized in *C. elegans*. First, efficient mobilization of *Mos1* is

\* Present address: Biologie Cellulaire de la Synapse, INSERM U. 497, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France.

possible in somatic cells. Second, heritable insertions of the transposon can be generated in the germ line. Third, genes that have been mutated by insertion can be rapidly identified using inverse polymerase chain reaction. Fourth, these insertions can subsequently be remobilized to generate deletion and frameshift mutations by imperfect excision.

*Mos1* is a member of the *mariner*/*Tc1* family, which was initially identified in the fruitfly *Drosophila mauritiana*<sup>2</sup>. Like the other members of the *mariner*/*Tc1* family, *Mos1* contains a single open reading frame (ORF), which encodes the transposase. The transposase binds to and cleaves at the inverted terminal repeats that are present at each end of the transposon (reviewed in refs 3 and 4). The *Mos1* transposase is the only protein necessary for transposition *in vitro*<sup>5</sup>. Because no additional factors are required for transposition, the *Mos1* transposon is capable of transposition in heterologous species *in vivo*<sup>6,7</sup>.

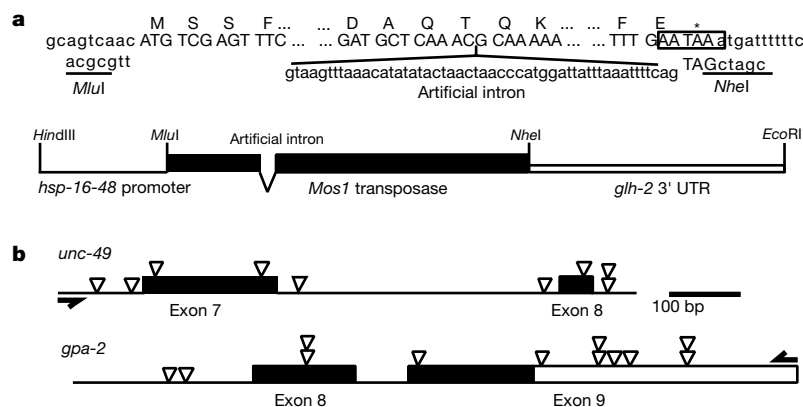
To determine whether the *Mos1* element could be mobilized in *C. elegans* cells we analysed *Mos1* transposition in somatic cells. We generated an extrachromosomal array expressing the *Mos1* transposase under the control of a heat-shock promoter (Fig. 1a). Another extrachromosomal array was generated using the *Mos1* transposon; this array also contained the dominant genetic marker *rol-6(sd)*, which causes animals to roll. These two strains were crossed and progeny carrying both arrays were subjected to heat shock as young adults. After 12 h the heat-shocked animals were collected, and *Mos1* insertions into arbitrary genes (*gpa-2* and *unc-49*) were assayed using polymerase chain reaction (PCR)<sup>8</sup>. For all insertions, the *Mos1* inverted terminal repeats were complete, the *Drosophila* sequences that flanked *Mos1* in the donor plasmid were no longer present, and the insertions all took place at a TA di-nucleotide. Transposon insertions were distributed uniformly in exons, introns and 3' non-coding sequences of *gpa-2* and *unc-49* (Fig. 1b). Thus, *Mos1* transposase was sufficient to catalyse insertion of *Mos1* into *C. elegans* chromosomes without *Drosophila* host factors.

To express the *Mos1* transposase in the germ line, we generated an extrachromosomal array expressing the transposase under the control of the *glh-2* promoter, which is expressed in the germ line<sup>9</sup>. Transposase expression in the germ line was determined by assaying for excision of transposons from a defined chromosomal location. Specifically, the *Mos1*-containing array was integrated on chromosome V to generate *oxIs25[mos1; rol-6(sd)]*. Heterozygous *oxIs25/dpy-11* worms segregated Dpy and Rol progeny as expected for these closely linked markers (Fig. 2). However, when the *glh-2::Transposase* transgene was crossed into this strain, 21% of the transposon arrays experienced loss of the transposons and

interspersed copies of *rol-6* (see Methods for calculation). Presumably, this catastrophic loss of the transposons and *rol-6* was caused by excision of the *Mos1* elements from the chromosome. A small fraction of these broken chromosomes recombined with the intact homologue. Specifically, 3% of the *dpy*-marked chromosomes (6 out of 198) also contained the integrated array. This hotspot for recombination is likely to arise as a result of double-strand breaks caused by *Mos1* excision<sup>10</sup>. Most visible excisions, however, were generated by intramolecular healing of the chromosome, which deleted the *Mos1* array. This conclusion was confirmed by analysing excision from a homozygous array: 37% of chromosomes from an *oxIs24[Mos1; rol-6(sd)]* homozygote lost the array when exposed to the transposase. In the homozygote, no wild-type chromosome was available for gene conversion or recombination.

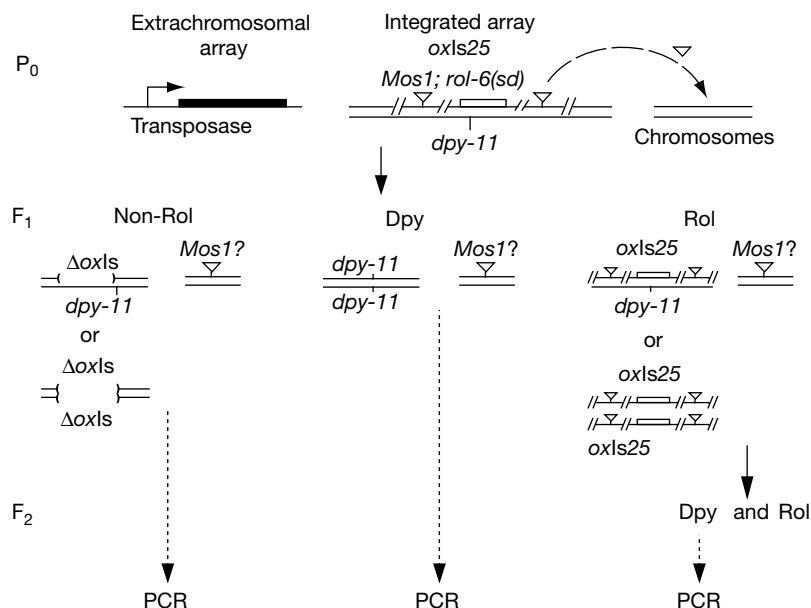
To determine if excision of *Mos1* from the array was associated with insertion in the genome, we screened for *de novo* insertions in animals that lacked the source array (Dpy progeny) or in individuals that had excised the array (non-Rol progeny). Insertions were identified in 1% of non-Rol progeny (2 out of 227) and in 10% of Dpy progeny (11 out of 116, Table 1). These results demonstrate that transposition of *Mos1* could be achieved in the *C. elegans* germ line. However, we observed that high rates of excision were not accompanied by high rates of insertion; these *in vivo* results are in agreement with previous biochemical data suggesting that these two processes are not mechanistically coupled<sup>11</sup>. We also tested whether *Mos1* could be mobilized from an extrachromosomal array into the chromosomes. The non-Rol progeny from double transgenic animals (*oxEx167[glh-2::Transposase]; oxEx164[Mos1; rol-6(sd)]*) were analysed for transposition events using PCR. Again, we detected an insertion frequency of 1% (3 insertions out of 302 progeny). Thus, these results closely match those obtained for integrated arrays.

We also determined whether the heat-shock promoter could mobilize the transposon from the integrated array in the chromosome. Animals bearing the transposase construct and the integrated transposon (*oxIs25[dpy-11; oxEx166[hsp::Transposase]]*) were subjected to heat-shock treatment and the Dpy progeny were analysed for transposition events. The heat-shock construct caused the efficient insertion of *Mos1* into new locations in the genome. Approximately 27% (25 out of 94) of F<sub>1</sub> or F<sub>2</sub> Dpy worms carried new transposon insertions (Table 1). F<sub>1</sub> progeny were also analysed for catastrophic excision, that is, for the appearance of non-Rol, non-Dpy progeny (Fig. 2). In contrast to results using the *glh-2* expression vector, catastrophic excision was not observed (Table 1). The absence of catastrophic excision is probably due to efficient



**Figure 1** Mobilization of *Mos1* in *C. elegans* somatic cells. **a**, Engineering of the *Mos1* transposase encoding sequence. Restriction sites were generated at the 5' and 3' ends of the coding sequence (new sequence is indicated under the original sequence). The endogenous polyadenylation signal (boxed) was disrupted; an artificial intron was introduced in the coding sequence to improve transposase expression<sup>24</sup>. Upper case

denotes the coding region. **b**, Localization of *Mos1* insertions in *unc-49* and *gpa-2* genes after induction of *Mos1* transposase expression in somatic cells. White triangles, insertion sites; black boxes, coding exons; white boxes, non-coding exonic sequence. Arrows indicate genomic primers used to amplify the insertions.



**Figure 2** Germline mobilization of *Mos1*. *Mos* transposase was expressed from an extrachromosomal array using either a *glh-2* or a heat-shock promoter. The *Mos1* transposon was contained in an array integrated on chromosome V (*oxls25*;*Mos1*;

*rol-6(sd)*). The array-containing chromosome was balanced by the *dpy-11*(*e224*) mutation. In the next generation, catastrophic excision of the transgene was observed (indicated as  $\Delta$ *oxls*) among the progeny.

template-dependent repair, in which the double-strand break is repaired by copying the original sequence from a sister chromatid. Catastrophic excision might be observed when using the *glh-2* promoter because of the timing of the excision event. For example, if excision precedes replication of the DNA then in the absence of a sister strand the break might be repaired by joining of the broken ends.

We next tested whether transposition could occur from an extrachromosomal array using the heat-shock promoter to express the transposase. Hermaphrodites bearing both arrays (*oxEx166* [*hsp::Transposase*]; *oxEx164* [*Mos1*; *rol-6(sd)*]) were subjected to heat-shock treatment as young adults. The non-Rol progeny were analysed by PCR for transposition events. New insertions were observed in 8.9% of the F<sub>1</sub> progeny (33 out of 369). The frequency of transposition was low but one of the principal limiting factors may have been the stability of the extrachromosomal array that was used as a transposon source. The initial experiments were performed when the array was only 20% stable and transposition frequencies were in the range of 5%; after many generations the array became more stable (75%) and transposition frequency reached 44%.

What is the distribution of insertions? To determine the location of the mobilized transposons, we cloned the left junctions of 58 insertions using inverse PCR. Insertion sites were distributed on all

six chromosomes (Fig. 3a), 11 were located in predicted exons and 10 in predicted introns. Sequences flanking both sides of the transposon were determined for ten of the localized insertions. In each case, the inverted terminal repeats were complete and flanked by a TA dinucleotide that arose from the duplication of the original TA found in the genomic sequence (Fig. 3b). Comparison of the 58 insertion-site sequences did not reveal a strong consensus motif for the target DNA, except for a TA-rich bias (72%) compared with the overall TA content of the genome (64%). Although exonic DNA comprises 27% of the genome<sup>12</sup>, only 19% of the insertions were found in predicted exons. Insertions into non-coding regions are probably enhanced by the higher TA content of these regions<sup>13,14</sup>.

The transposition events documented above were all excisions from an array of transposons residing in *Drosophila* DNA. To determine whether the transposase acts on a single *Mos1* transposon in a *C. elegans* chromosome, we remobilized *oxTi4* (Transposon insertion) using the *glh-2::Transposase* construct. A variety of excision events were detected, including the three-nucleotide excision footprint previously characterized for *Mos1* excisions<sup>15</sup>, as well as smaller footprints, deletions and even incomplete excisions (Fig. 3c). As these products could arise from excision events in somatic cells, we identified a strain in which the transposon had excised in the

**Table 1 Comparison of excision and insertion frequencies**

| Expression vector | Experiment | F <sub>1</sub> non-Rol                               |        |                    |      | F <sub>1</sub> Dpy                         |    |                                |     | F <sub>2</sub> Dpy                         |     |                                |     |
|-------------------|------------|------------------------------------------------------|--------|--------------------|------|--------------------------------------------|----|--------------------------------|-----|--------------------------------------------|-----|--------------------------------|-----|
|                   |            | Recombinants or excisions (non-Rols/Rols + non-Rols) |        | Insertions/non-Rol |      | Recombinants ( <i>dpy-11</i> <i>oxls</i> ) |    | Insertions/F <sub>1</sub> Dpys |     | Recombinants ( <i>dpy-11</i> <i>oxls</i> ) |     | Insertions/F <sub>2</sub> Dpys |     |
| <i>glh-2</i>      |            | 3/691                                                | 0.43%  | ND                 |      | 0/92                                       | 0% | ND                             |     | ND                                         |     | ND                             |     |
|                   | 1          | 541/4,078                                            | 13.3%  | 2/188              | 1.1% |                                            |    | ND                             |     |                                            |     | ND                             |     |
|                   | 2          | 138/651                                              | 21.2%  | ND                 |      | 0/39                                       | 0% | 2/39                           | 5%  | 2/19                                       | 10% | 2/17                           | 11% |
|                   | 3          | 250/1,191                                            | 20.8%  | 0/39               | 0%   | 2/37                                       | 5% | 0/35                           | 0%  | 2/27                                       | 7%  | 7/25                           | 28% |
|                   | Total      | 929/5,920                                            | 15.7%  | 2/227              | 0.9% | 2/76                                       | 3% | 2/74                           | 3%  | 4/46                                       | 9%  | 9/42                           | 21% |
| <i>hsp</i>        |            | 4/1,048                                              | 0.38%* | ND                 |      |                                            |    | ND                             |     | 4/24                                       | 17% | 6/24                           | 25% |
|                   | 1          |                                                      |        | ND                 |      | 0/13                                       | 0% | 0/13                           | 0%  | 0/17                                       | 0%  | 4/17                           | 23% |
|                   | 2          | 0/41                                                 | 0%*    | ND                 |      | 1/20                                       | 5% | 7/20                           | 35% | 3/20                                       | 15% | 8/20                           | 40% |
|                   | 3          | 1/140                                                | 0.71%* | ND                 |      |                                            |    |                                |     |                                            |     |                                |     |
|                   | Total      | 5/1,229                                              | 0.41%* | ND                 |      | 1/33                                       | 3% | 7/33                           | 21% | 7/61                                       | 11% | 18/61                          | 30% |

*glh-2* and heat-shock (*hsp*) promoters were used to drive *Mos* transposase expression in the germ line. New *Mos1* insertions were identified by PCR. Recombinants between *dpy-11* and the insertion appeared as non-Rol progeny or as Dpy worms also containing *Mos1* flanked by original *Drosophila* genomic sequences (compare with row marked 'none'). Excisions were recognized as the number of non-Rol progeny appearing in the F<sub>1</sub> greater than expected for recombination. ND, not done.

\* Number on non-Rol progeny does not significantly differ from control lacking a transposase-expressing construct (3/691, 0.43%).



germ line. The excision left a 3-base pair (bp) footprint and the duplicated TA dinucleotide, which together resulted in a +2 frame-shift. These data indicate that single copies of the *Mos1* *Drosophila* transposon can excise from *C. elegans* DNA in the germ line to introduce frameshift or deletion mutations at the transposon insertion site.

To determine the mutagenic rate of *Mos1*, we performed an  $F_2$  selection for animals defective in the detection of high osmolarity (Osm). Early adult worms, which contained both the *Mos1* transposon and transposase arrays (*oxEx166[hsp::Transposase]*; *oxEx229[Mos1]*), were subjected to heat-shock treatment, and  $F_2$  worms were tested for the Osm phenotype. A total of 161,064 haploid genomes were screened and ten Osm mutants were identified. To compare the *Mos1* mutagenic rate to that of a chemical mutagen, we screened for Osm mutants using the mutagen *N*-nitroso-*N*-ethylurea (ENU). In this screen 1,936 haploid genomes

were screened and eight strong Osm mutants were identified. From this data *Mos1* seems to be approximately 60-fold less efficient as a mutagen than ENU. Six out of the ten Osm mutants were associated with a *Mos1* insertion. After out-crossing the strains many times, the relevant gene containing the *Mos1* insertion was identified using inverse PCR and the mutation confirmed by complementation or mapping experiments. The other four mutants seem to be 'hit and run' events in which a transposon was no longer detected in the mutant strain. It is possible that the hit and run events were caused by the mobilization of other transposons in the genome. There are at least 55 copies of a Mariner-like element (MLE), which is closely related to *Mos1* (refs 16, 17) in the *C. elegans* haploid genome. However, no changes in MLE distribution were detected in eight strains in which *Mos1* insertions had occurred (Southern blot not shown).

All of the insertion mutations were caused by insertions into exons. Insertions into introns are not likely to be mutagenic, as they can be spliced out during transcription. To improve mutagenicity we constructed a transposon that would disrupt transcription. A 378-bp fragment containing a consensus splice acceptor site followed by the *unc-54* polyadenylation signal sequence was inserted into the *SalI* site of *Mos1* (*MosPolyA*), and extrachromosomal arrays were generated. The recombinant transposon was mobilized using the integrated heat-shock construct (*oxIs31*). Progeny were assayed for insertions by single-worm PCR and the location of the insert was determined by inverse PCR. Insertions were generated from the wild-type and *MosPolyA* elements at identical rates (6 out of 108  $F_1$  progeny for each, Fig. 3a). These data also show that short, single-copy DNA can be introduced into the *C. elegans* genome using the *Mos1* transposon. However, a 2-kb insertion decreased *Mos* transposition frequency by at least 10-fold (0 insertions from 212 progeny).

These experiments demonstrate the feasibility of using a heterologous transposon as a mutagen in *C. elegans*. The use of *Mos* should significantly accelerate forward genetic strategies. Furthermore, mobilization of an engineered transposon containing, for example, recombinase or meganuclease sites should enable the development of new tools to manipulate the *C. elegans* genome. □

## Methods

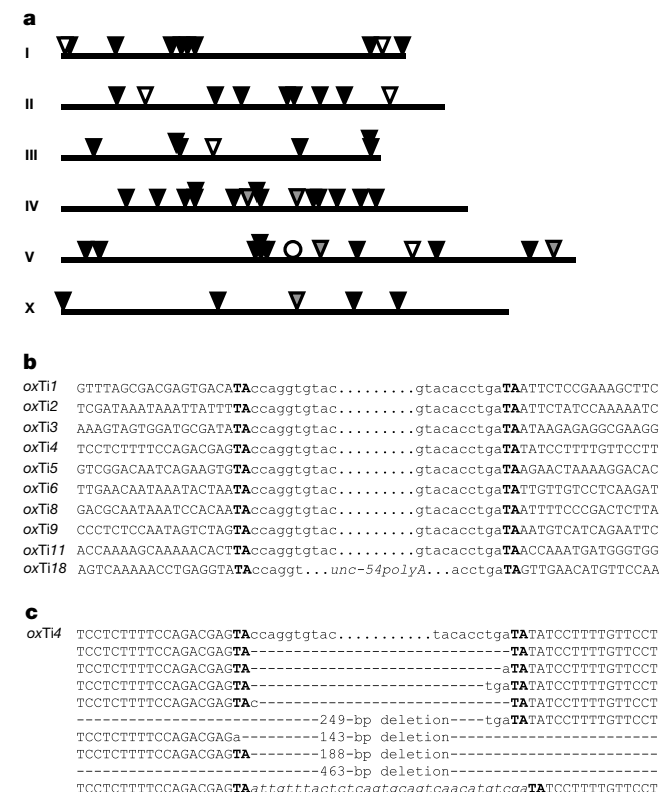
### Construction of transgenic strains

To build the *hsp::Transposase* plasmid, the transposase-encoding sequence was amplified from pBluescribe M13+/+ (*ref. 18*), modified as described in Fig. 1 and subcloned between the *hsp-16-48* promoter<sup>8,19</sup> and the *glh-2* 3' untranslated region (residues 35,383–36,190 of cosmid C55B7)<sup>9</sup>. In the *glh-2::Transposase* construct, the transposase ORF was preceded by a 2.2-kb *glh-2* promoter fragment (residues 29,882–32,095 of cosmid C55B7). To generate the *oxEx166[hsp::Transposase]* array, the gonads of *lin-15(n765)* hermaphrodites were injected with the gel-purified fragments of the following constructs: *hsp::Transposase* (injection concentration of 10 ng  $\mu\text{L}^{-1}$ ); the *unc-122::gfp* construct (pPD97/98) that expresses green fluorescent protein (GFP) in the coelomocytes (gift of P. Sengupta) (5 ng  $\mu\text{L}^{-1}$ ); and EKL15(*lin-15+*)<sup>20</sup> (10 ng  $\mu\text{L}^{-1}$ ). *oxEx167[glh-2::Transposase]* was generated by injecting the *glh-2::Transposase* construct (10 ng  $\mu\text{L}^{-1}$ ) with *lin-15(+)* and *unc-122::gfp* fragments. The *Mos1*-containing array *oxEx164[Mos1; rol-6(sd)]* was built by injecting a fragment containing the 1.3-kb *Mos1* element flanked by *Drosophila simulans* sequences<sup>18</sup> (10 ng  $\mu\text{L}^{-1}$ ) and a fragment of pRF4 (*ref. 21*) (*rol-6(sd)*) (10 ng  $\mu\text{L}^{-1}$ ). In each case, plasmid backbones were removed from purified fragments and wild-type worm genomic DNA that was digested with *EcoRV* was co-injected to optimize germline expression (final DNA concentration of 100 ng  $\mu\text{L}^{-1}$ )<sup>22</sup>. *oxEx229[Mos1]* was generated by injecting uncut *Mos1* plasmid<sup>18</sup> at 10 ng  $\mu\text{L}^{-1}$ , *myo-2::GFP* marker (pPD118.33) at 2 ng  $\mu\text{L}^{-1}$  and pBluescriptKII at 88 ng  $\mu\text{L}^{-1}$ .

### Detection of *Mos1* transposition

For detection of somatic transposition, DNA purification and PCR were performed as described<sup>12</sup>. *Mos1* insertions were detected at high frequency ( $2.5 \pm 1.0$  inserts in 10 ng genomic DNA, mean  $\pm$  s.d.;  $n = 5$  experiments). Given that the maximum distance of the inserts from our gene primers was approximately 1 kb, we estimated that an average of ten insertions occurred per cell in heat-shocked animals.

The integrated array *oxIs25[Mos1; rol-6(sd)]* contained 15–20 *Mos1* elements as determined by Southern blot analysis. Catastrophic excision of the array was analysed by PCR using one primer complementary to *Mos1* (oJL102: 5'-CAACCTTGACTGTGCGAACACCATAG-3') and one primer complementary to *D. simulans* flanking DNA (oJL104: 5'-ACAAAGAGCGAAGCGACGAGT-3'). The probability of a single chromosome



**Figure 3** *Mos1* genomic insertions. **a**, Distribution of *Mos1* inserts on the physical map of the *C. elegans* genome. Black triangles, insertions from an extrachromosomal array, white triangles, insertions from the integrated array *oxIs25*; grey triangles, insertions of *MosPolyA*; white circle, position of *oxIs25*, the integrated array of *Mos1* transposons. Insertion sites can be obtained from Wormbase (<http://www.wormbase.org/>). **b**, DNA sequence of *Mos1* *de novo* insertions. Genomic fragments that flank the transposon left end were isolated by inverse PCR and sequenced. A primer was designed in the genomic region to the right of the insert and used with a *Mos1*-specific primer to amplify and sequence the right end flanking the fragment. The insertion *oxTi18* was amplified using worm-specific primers and the entire element was sequenced. At insertion sites TA, dinucleotides (bold) were duplicated during the process of transposon integration. Lower case, *Mos1* sequence; upper case, genomic sequence. **c**, Lesions generated by excision of the *oxTi4* insert. The extrachromosomal transgene [*glh-2::Transposase*] was crossed into animals that were homozygous for the *oxTi4* insertion. PCR was used to analyse the *oxTi4* insertion site after the loss of *Mos1*. Top line is the sequence of *oxTi4*. Lower case, *Mos1* sequence; upper case, genomic sequence; bold, TA dinucleotide duplicated during *Mos1* insertion; dash, deleted base pairs. The bottom lines are the excision products. The fourth line represents the footprint obtained from the germline excision event. The insertion (bottom line, italic letters) corresponds to an internal fragment of *Mos1* (residues 147–178).

containing the array of *Mos1* elements experiencing catastrophic excision could be derived from a Punnett square where the ratio, *R*, of non-Rol worms over the probability of recombination, *p*, is  $R = 1/4p + 1/4p + (1/2p)^2$ .

To detect *de novo* insertions, the presence of *Mos1* was detected by PCR using two primers located in the transposon (oJL102 and oJL103: 5'-TCTGCGAGTTGTTTT GCGTTTGTAG-3'). The absence of *D. simulans* flanking sequence was checked using oJL103 and oJL104. Furthermore, we performed a positive control for PCR on each DNA sample using primers from the *cha-1* gene.

Insertions of *Mos1* were localized in the genome by inverse PCR: 100 ng total genomic DNA was digested with *Sau3A*, self-ligated under dilute conditions, and then subjected to two rounds of nested PCR using the following primers: oJL103/oJL114 5'-AAAGATTCA GAAGTGGGTAGATGGG-3' (first PCR), oJL115 5'-GCTCAATTGCGGCCAACT ATG-3' / oJL116 5'-GAACGAGAGGCAGTGGAGAGG-3' (second PCR). PCR products were sequenced using oJL115 as a primer.

To build the recombinant element *MosPolyA*, the *unc-54* polyadenylation site was amplified by PCR using oJL125 (5'-AATATTTAAATTTTCAGCATC-3') and oJL126 (5'-CGTAACATATGATAAGGTATTT-3') and the resulting 378-bp fragment was ligated into the *Sall* site of *Mos1*. Transposons from multiple arrays of wild-type (oEx338, oEx339, oEx340) and recombinant *Mos* elements (oEx344, oEx345, oEx346) were mobilized in parallel. The presence of *MosPolyA* in strains was detected using oJL102 and oJL103 in single-worm PCR reactions. Insertion sites of *MosPolyA* were determined by inverse single-worm PCR. Five worms were lysed in 20 µl lysis buffer<sup>23</sup> and 0.625 worm equivalents were digested with *HhaI* or *MseI*, then ligated under dilute conditions. Nested PCR reactions were performed using primer 1A (5'-GACCTTGTA AGTGT-CAACCTTGACTG-3')/1B (5'-GACAATCGATAAATATTACGTTTGCGAGAC-3') in the first reaction and oJL102/2B (5'-CATCTATATGTTGCAACCGACATTTCCC-3') in the second reaction. Amplified products were gel-purified and sequenced using primer 2B.

## Characterization of *Mos1* re-excision events

*Mos1* was remobilized from oxiT4 (insertion in chromosome II at position 4,948 in cosmid T13C25). DNA from the progeny of oxiT4; oxiEx167[*glh-2::Transposase*] hermaphrodites was prepared. A first PCR round was performed with primers located 1,671 nucleotides upstream and 3,144 bp downstream to oxiT4 (oJL149 5'-AAGTATGGCCAAACGACCCG ACAC-3' and oJL150 5'-GCATTGGCACCTTTCTCCCTTCT-3', respectively). Short PCR fragments were detected in 18 out of 27 cultures. A second round was performed using nested primers (oJL145 5'-ACAGGCAGCATTTGTAGTCT-3' and oJL148 5'-AGGCTGCCTCGTAAGTTCCTACAG-3', respectively). Short PCR products were gel-purified, subcloned and sequenced.

To clone worms that lost the oxiT4 insertion, pools of 15 individuals from oxiT4; oxiEx167[*glh-2::Transposase*] progeny that lost the transposase array were transferred to fresh plates and allowed to lay eggs for 24 h. We then analysed adult worms by a single round of PCR using the primers oJL145–oJL148. Sixty individuals were cloned from the progeny of the pool exhibiting short PCR product, and analysed at the next generation to identify clones that lost oxiT4. One inherited excision was identified from among 954 progeny tested.

## Osm screen

Early adult worms, which contained both the *Mos1* transposon and transposase arrays (oEx166[*hsp::transposase*]; oxiEx229[*Mos1*]), were subjected to heat-shock treatment at 35°C for 1 h. F<sub>2</sub> progeny were placed in a ring of 4 M fructose. Wild-type worms remain in the ring of fructose, whereas mutants that are unable to detect the high osmolarity cross the border.

Received 6 April; accepted 16 July 2001.

- Plasterk, R. H. A. & van Luenen, H. G. A. M. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 97–116 (Cold Spring Harbor Laboratory, New York, 1997).
- Jacobson, J. W., Medhora, M. M. & Hartl, D. L. Molecular structure of a somatically unstable transposable element in *Drosophila*. *Proc. Natl Acad. Sci. USA* **83**, 8684–8688 (1986).
- Hartl, D. L., Lohe, A. R. & Lozovskaya, E. R. Modern thoughts on an ancient mariner: function, evolution, regulation. *Annu. Rev. Genet.* **31**, 337–358 (1997).
- Plasterk, R. H., Izsvak, Z. & Ivics, Z. Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* **15**, 326–332 (1999).
- Tosi, L. R. & Beverley, S. M. *cis* and *trans* factors affecting *Mos1* mariner evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Res.* **28**, 784–790 (2000).
- Gueiros-Filho, F. J. & Beverley, S. M. Trans-kingdom transposition of the *Drosophila* element mariner within the protozoan *Leishmania*. *Science* **276**, 1716–1719 (1997).
- Coates, C. J., Jasinskiene, N., Miyashiro, L. & James, A. A. Mariner transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl Acad. Sci. USA* **95**, 3748–3751 (1998).
- van Luenen, H. G., Colloms, S. D. & Plasterk, R. H. Mobilization of quiet, endogenous Tc3 transposons of *Caenorhabditis elegans* by forced expression of Tc3 transposase. *EMBO J.* **12**, 2513–2520 (1993).
- Gruidl, M. E. *et al.* Multiple potential germ-line helicases are components of the germ-line-specific P granules of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 13837–13842 (1996).
- Lohe, A. R., Timmons, C., Beerman, I., Lozovskaya, E. R. & Hartl, D. L. Self-inflicted wounds, template-directed gap repair and a recombination hotspot. Effects of the mariner transposase. *Genetics* **154**, 647–656 (2000).
- van Luenen, H. G., Colloms, S. D. & Plasterk, R. H. The mechanism of transposition of Tc3 in *C. elegans*. *Cell* **79**, 293–301 (1994).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Sulston, J. E. & Brenner, S. The DNA of *Caenorhabditis elegans*. *Genetics* **77**, 95–104 (1974).
- Waterston, R. H., Sulston, J. E. & Coulson, A. R. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T.,

Meyer, B. J. & Priess, J. R.) 23–45 (Cold Spring Harbor Laboratory, New York, 1997).

- Bryan, G., Garza, D. & Hartl, D. Insertion and excision of the transposable element mariner in *Drosophila*. *Genetics* **125**, 103–114 (1990).
- Sedensky, M. M., Hudson, S. J., Everson, B. & Morgan, P. G. Identification of a mariner-like repetitive sequence in *C. elegans*. *Nucleic Acids Res.* **22**, 1719–1723 (1994).
- Robertson, H. M. & Lampe, D. J. Recent horizontal transfer of a mariner transposable element among and between *Diptera* and *Neuroptera*. *Mol. Biol. Evol.* **12**, 850–862 (1995).
- Medhora, M., Maruyama, K. & Hartl, D. L. Molecular and functional analysis of the mariner mutator element *Mos1* in *Drosophila*. *Genetics* **128**, 311–318 (1991).
- Candido, E. P. *et al.* Structure, organization, and expression of the 16-kDa heat shock gene family of *Caenorhabditis elegans*. *Genome* **31**, 690–697 (1989).
- Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137**, 987–997 (1994).
- Kramer, J. M., French, R. P., Park, E. C. & Johnson, J. J. The *Caenorhabditis elegans* *rol-6* gene, which interacts with the *sqt-1* collagen gene to determine organismal morphology, encodes a collagen. *Mol. Cell. Biol.* **10**, 2081–2089 (1990).
- Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct requirements for somatic and germline expression of a generally expressed *Caenorhabditis elegans* gene. *Genetics* **146**, 227–238 (1997).
- Williams, B. D., Schrank, B., Huynh, C., Shownkeen, R. & Waterston, R. H. A genetic mapping system in *Caenorhabditis elegans* based on polymorphic sequence-tagged sites. *Genetics* **131**, 609–624 (1992).
- Fire, A., Harrison, S. W. & Dixon, D. A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189–198 (1990).

## Acknowledgements

We thank D. Hartl, H. van Leunen, R. Plasterk, S. Beverley and K. Bennett for gifts of DNA constructs, and O. Hobert for unpublished information. We also thank H. Rausch for initial studies using Tc3 and C. Johnston, E. Lyman and M. Miller for screening help. This work was funded by an NSF grant. J.-L.B. was supported by INSERM, EMBO and the Fondation Simone et Cino del Duca.

Correspondence and requests for materials should be addressed to E.M.J. (e-mail: jorgensen@biology.utah.edu).

# Role of thrombin signalling in platelets in haemostasis and thrombosis

Gilberto R. Sambrano\*†, Ethan J. Weiss\*†, Yao-Wu Zheng\*†, Wei Huang\* & Shaun R. Coughlin\*‡

\* Cardiovascular Research Institute and ‡ Departments of Medicine and of Cellular and Molecular Pharmacology, University of California, 513 Parnassus Avenue, San Francisco, California 94143-0130, USA

† These authors contributed equally to this work

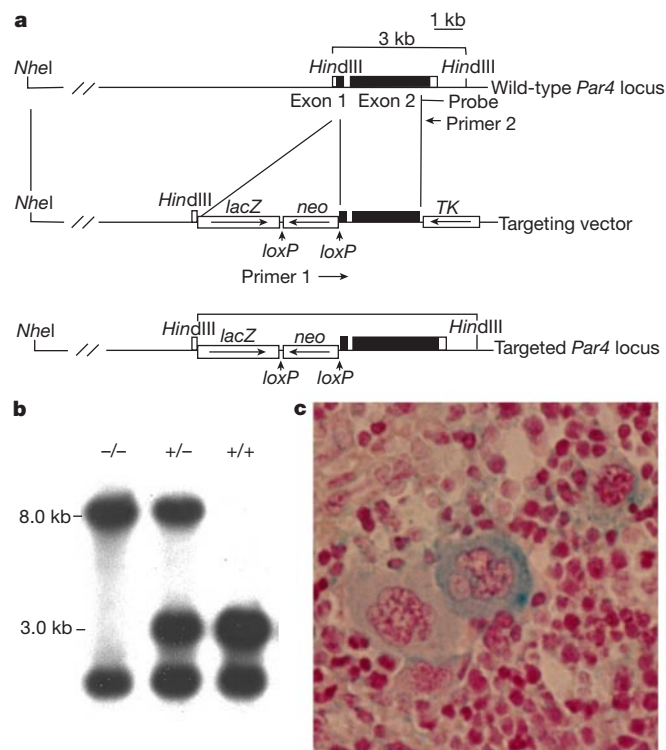
Platelets are critical in haemostasis and in arterial thrombosis, which causes heart attacks and other events triggered by abnormal clotting<sup>1–5</sup>. The coagulation protease thrombin is a potent activator of platelets *ex vivo*<sup>6</sup>. However, because thrombin also mediates fibrin deposition and because multiple agonists can trigger platelet activation<sup>7</sup>, the relative importance of platelet activation by thrombin in haemostasis and thrombosis is unknown. Thrombin triggers cellular responses at least in part through protease-activated receptors (PARs)<sup>8</sup>. Mouse platelets express PAR3 and PAR4 (ref. 9). Here we show that platelets from PAR4-deficient mice failed to change shape, mobilize calcium, secrete ATP or aggregate in response to thrombin. This result demonstrates that PAR signalling is necessary for mouse platelet activation by thrombin and supports the model that mouse PAR3 (mPAR3) does not by itself mediate transmembrane signalling but instead acts as a cofactor for thrombin cleavage and activation of mPAR4 (ref. 10). Importantly, PAR4-deficient mice had markedly prolonged bleeding times and were protected in a model of arteriolar thrombosis. Thus platelet activation by thrombin is necessary for normal haemostasis and may be an important target in the treatment of thrombosis.

We disrupted the *Par4* gene in embryonic stem cells by inserting the *lacZ* coding sequence into exon 1 such that β-galactosidase

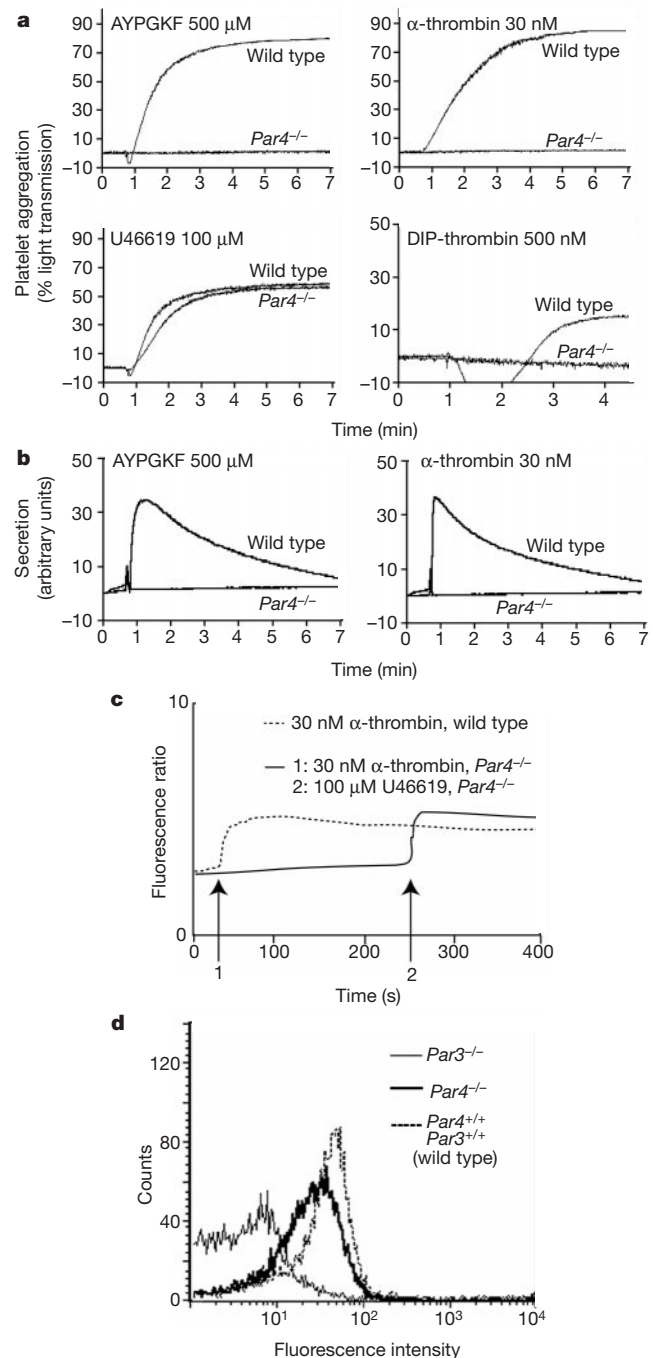
would be expressed in place of PAR4, and mice heterozygous for this allele were generated (Methods and Fig. 1).  $\beta$ -galactosidase staining of  $Par4^{+/-}$  embryonic liver and adult bone marrow and spleen was restricted to megakaryocytes (Fig. 1c and data not shown), consistent with previous *in situ* hybridization for  $Par4$  messenger RNA<sup>9</sup>. Polymerase chain reaction with reverse transcription (RT-PCR) and northern analysis of spleen RNA for PAR4 exon 2 sequence confirmed the loss of PAR4 mRNA expression (not shown), and platelets from  $Par4^{-/-}$  mice were unresponsive to the PAR4-agonist peptide AYPGKF, confirming loss of PAR4 function (Fig. 2a).  $Par4^{-/-}$  offspring were obtained at the expected mendelian frequency from  $Par4$  heterozygote crosses ( $Par4^{+/+}$ , 29%;  $Par4^{+/-}$ , 46%; and  $Par4^{-/-}$ , 25%;  $n = 275$ ) and were indistinguishable from wild-type littermates in appearance and size. Mass at weaning (mean  $\pm$  s.d.) was  $11.0 \pm 2.1$  g ( $n = 21$ ) and  $10.6 \pm 1.8$  g ( $n = 21$ ) for wild-type and  $Par4^{-/-}$  mice, respectively. Male and female  $Par4^{-/-}$  mice were fertile, and females carried pregnancies to term without evident bleeding. Haematocrits were  $48.5 \pm 1.1\%$  in wild-type and  $46.5 \pm 3.0\%$  in  $Par4^{-/-}$  mice ( $n = 5$ ). Erythrocyte, leukocyte and platelet counts in wild-type and  $Par4^{-/-}$  mice were not different. Transmission electron microscopic examination of wild-type and  $Par4^{-/-}$  platelets revealed no apparent differences.

Remarkably,  $Par4^{-/-}$  platelets were unresponsive to thrombin (Fig. 2a–c). Thrombin concentrations of 30, 60 and even 500 nM

failed to trigger shape change, ATP secretion, aggregation or increases in cytoplasmic calcium in  $Par4^{-/-}$  platelets. In wild-type mouse platelets, the half-maximal effective concentration ( $EC_{50}$ ) for these responses was 1–3 nM and maximal responses were achieved at 10–30 nM thrombin (Fig. 2, ref. 9 and data not shown).



**Figure 1** Disruption of the mouse PAR4 gene. **a**, Targeting construct. Exon 1 encodes a 5' untranslated sequence, start codon and signal peptide. Exon 2 encodes the remainder of the receptor protein and a 3' untranslated sequence<sup>26</sup>. A  $\beta$ -galactosidase ('lacZ') complementary DNA and a *neo* cassette bound by tandem *loxP* sites were inserted such that the start codon of  $\beta$ -galactosidase supplanted that of PAR4. Large arrows indicate transcriptional orientation. Homologous recombination in embryonic stem cells was confirmed by Southern analysis and by PCR across the short arm. **b**, Southern analysis of mice generated from targeted embryonic stem cells with the probe indicated in **a**, which was external to the targeting construct. The probe detected 3-kb and 8-kb *HindIII* fragments representing the wild-type and targeted PAR4 alleles, respectively. **c**,  $\beta$ -galactosidase expression in embryonic liver. LacZ staining (blue) of haematopoietic tissues from *neo*-excised  $Par4^{+/-}$  mice confirmed megakaryocyte-specific expression. Adult liver or spleen yielded similar results and wild-type controls showed no staining.



**Figure 2**  $Par4^{-/-}$  platelets lack thrombin responses. Washed platelets from wild-type ( $Par4^{+/+}$ ) or  $Par4^{-/-}$  mice were examined by lumiaggregometry for agonist-triggered aggregation (**a**) or ATP secretion (**b**) or were loaded with Fura2 and examined for agonist-triggered increases in cytosolic calcium measured fluorometrically (**c**). The PAR4-agonist AYPGKF (500  $\mu$ M),  $\alpha$ -thrombin (30 nM), DIP-thrombin (500 nM) or the thromboxane-receptor-agonist U46619 (100  $\mu$ M) were added at time  $t = 45$  s in **a** and **b** and when indicated by the arrow in **c**. These experiments were replicated 3–6 times. No secretion, aggregation or calcium mobilization was seen in  $Par4^{-/-}$  platelets even at 500 nM  $\alpha$ -thrombin. **d**, Confirmation of PAR3 protein expression on the surface of  $Par4^{-/-}$  platelets by flow cytometry. Northern analysis of spleen mRNA from wild-type and  $Par4^{-/-}$  mice showed no difference in PAR3 mRNA levels.



Responses to the thromboxane-receptor-agonist U46619 and to a peptide agonist for the collagen receptor GPVI (ref. 11) were normal in *Par4*<sup>-/-</sup> platelets (Fig. 2 and data not shown). Thus PAR4 is necessary for thrombin responses in mouse platelets.

Like PAR4, PAR3 is expressed by mouse platelets and by megakaryocytes in mouse spleen and bone marrow<sup>12</sup>. Northern analysis of spleen mRNA from *Par4*<sup>-/-</sup> and wild-type mice revealed no difference in PAR3 mRNA levels (data not shown), and cell surface expression of mPAR3 was not significantly different in wild-type and *Par4*<sup>-/-</sup> mouse platelets (Fig. 2d). Thus PAR3 expression is not sufficient for platelet activation by thrombin. Deletion of the *Par3* gene does, however, decrease mouse platelet responsiveness to thrombin<sup>9</sup>. Also, in heterologous expression systems, mPAR3 does not by itself mediate thrombin signalling but instead serves as a cofactor that promotes cleavage and activation of PAR4 at low thrombin concentrations<sup>10</sup>. This model predicts that thrombin responses in mouse platelets should be dependent on PAR4. Thus our data strongly support a cofactor role for mPAR3 even in its native environment, the mouse platelet.

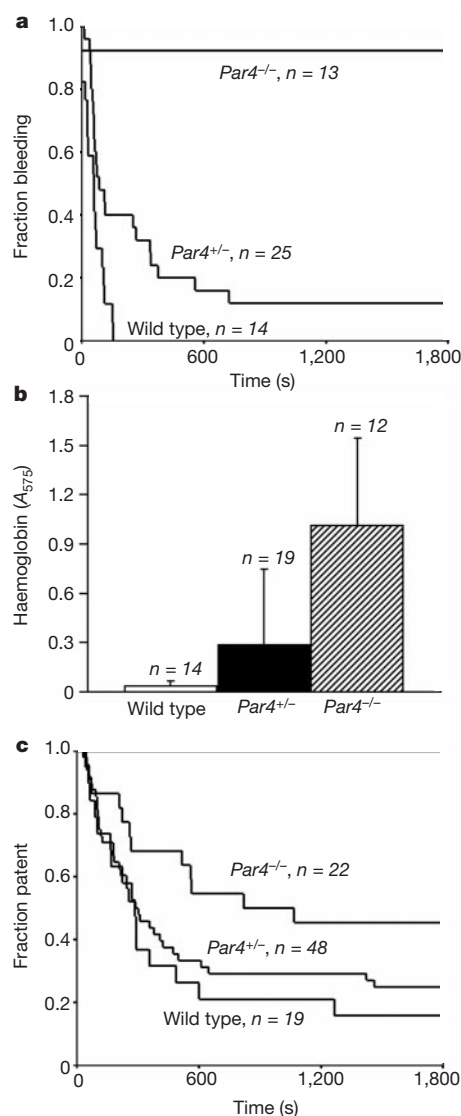
GPIb $\alpha$ , part of an abundant and multifunctional protein complex on the platelet surface, can bind thrombin<sup>13</sup>. GPV, a component of the GPIb complex, is a thrombin substrate<sup>13</sup>. Several lines of evidence suggest that the GPIb complex contributes, directly or indirectly, to platelet activation by thrombin<sup>13–15</sup>. Moreover, preparations of diisopropylfluorophosphate (DIP)-thrombin, a thrombin with a blocked active site, can activate platelets, and this activity is increased in platelets lacking GPV or in platelets in which GPV has been cleaved by active thrombin<sup>16</sup>. These studies have raised the possibility that non-proteolytic signalling mechanisms contribute to platelet activation by thrombin. High concentrations of DIP-thrombin (500 nM) did weakly activate wild-type mouse platelets (Fig. 2a). However, DIP-thrombin failed to elicit responses in *Par4*<sup>-/-</sup> platelets (Fig. 2a) even when platelets were pre-incubated with 100  $\mu$ M epinephrine to lower cyclic AMP or with  $\alpha$ -thrombin<sup>16</sup> to cleave GPV (data not shown).

At face value, the requirement for a PAR for platelet activation by DIP-thrombin suggests that PAR cleavage, perhaps by residual unblocked thrombin in DIP-thrombin preparations or by another protease in the system, is necessary for DIP-thrombin signalling. More broadly, the loss of thrombin-triggered calcium signalling, secretion and aggregation in *Par4*<sup>-/-</sup> platelets (Fig. 2a–c) suggests that the GPIb complex is not sufficient to mediate platelet responses to thrombin in the absence of PAR function. These results do not exclude the possibility that the GPIb complex contributes to platelet activation by thrombin. It is certainly possible that thrombin binding to GPIb serves a cofactor function for cleavage of platelet PARs. It is also possible that the experimental conditions used here (see Methods) mask a PAR-independent signalling pathway, that PAR4 is somehow necessary for a GPIb-mediated non-proteolytic signalling function of DIP-thrombin, and/or that GPIb interacts with or promotes PAR signalling via its interactions with downstream signalling complexes<sup>13</sup> or by unknown mechanisms.

The absence of thrombin responses in *Par4*<sup>-/-</sup> platelets provided a long-awaited opportunity to assess the importance of platelet activation by thrombin in haemostasis and thrombosis. Haemostasis was assessed by determining tail bleeding times. Median tail bleeding times were 1.1 and 1.5 min, respectively, for wild-type and *Par4*<sup>+/-</sup> mice, with more variability noted in the latter (Fig. 3a). By contrast, in *Par4*<sup>-/-</sup> mice, tail bleeding times were almost uniformly longer than 20 min, the arbitrary time for terminating the experiment (Fig. 3a). The quantity of blood lost during the bleeding-time assays roughly paralleled the bleeding times (Fig. 3b): *Par4*<sup>-/-</sup> mice lost 25 times more blood than wild-type mice. These results strongly indicate that activation of platelets by thrombin is necessary for normal haemostasis.

*Par4*<sup>-/-</sup> mice were protected in a mouse model of arteriolar thrombosis (Fig. 3c). Mouse mesenteric arterioles were injured by

exposure to ferric chloride, and time to thrombosis was measured by videomicroscopy<sup>17</sup>. This arteriolar thrombosis model displays platelet dependence and cyclic flow variations like arterial thrombosis models in other systems. In wild-type, *Par4*<sup>+/-</sup> and *Par4*<sup>-/-</sup> mice, median time to occlusion was 289, 292 and 822 s, respectively. Moreover, 30 min after injury, 45% of arterioles in *Par4*<sup>-/-</sup> mice remained open, compared with 16 and 25% in wild-type and *Par4*<sup>+/-</sup> mice, respectively ( $P < 0.04$ , chi-squared test). These data strongly indicate that platelet activation by thrombin is important in this model of arteriolar thrombosis and motivate future studies in models of arterial and venous thrombosis.



**Figure 3** Haemostasis and thrombosis in *Par4*<sup>-/-</sup> mice. **a**, Tail bleeding times were determined and the data expressed as the fraction still bleeding as a function of time. The experiment was terminated at 20 min. The fraction still bleeding at 20 min was different for each genotype ( $P < 0.001$ , chi-squared test). **b**, Blood loss (mean  $\pm$  s.d.) during the bleeding time assays in **a** was assessed by measuring the absorbance ( $A_{575}$ ) of haemoglobin accumulated in the saline in which tails were placed. Blood loss in wild-type (*Par4*<sup>+/+</sup>) mice was different from that in *Par4*<sup>-/-</sup> mice ( $P < 0.01$ , *t*-test with Bonferroni correction). **c**, Mesenteric arteriole thrombosis. Time from application of ferric chloride to cessation of flow was measured. Data are expressed as the fraction of arterioles still patent at 30 min. The fraction of arterioles still patent at 30 min was different between wild-type (16%) and *Par4*<sup>-/-</sup> (45%) mice ( $P < 0.04$ , chi-squared test). Log-rank analysis of wild-type and *Par4*<sup>-/-</sup> data in **a** and **c** also gave significant *P*-values.

Haemostatic and thrombotic responses are orchestrated by a complex interplay between platelets, plasma proteins, leukocytes and the vessel wall<sup>7,18</sup>. Thrombin itself is clearly important in these processes: in addition to activating platelets, thrombin mediates fibrin formation and has several other actions<sup>7,8,19</sup>. Similarly, platelets are clearly important in haemostasis and thrombosis, but many agonists that are present at sites of vascular injury can activate platelets<sup>13,20–22</sup>. Thus, identifying the molecular mechanisms by which thrombin activates platelets and defining the relative importance of thrombin signalling has been an important goal.

Our results provide strong genetic evidence that protease-activated receptors are necessary for platelet activation by thrombin, and that such signalling is important in haemostasis and thrombosis, despite the panoply of agonists and potentially redundant pathways involved in these processes. PAR-stimulated intracellular signalling pathways activate  $\alpha_{IIb}/\beta_3$  to directly trigger platelet aggregation, excite release of paracrine and autocrine mediators that amplify platelet activation, and trigger the display of procoagulant activity on the platelet surface, which promotes additional thrombin production and fibrin formation<sup>7,8,19</sup>. The effects of PAR4 deficiency on haemostasis and thrombosis might be explained by the loss of one or more of these functions. Human and mouse platelets do differ in the specific PAR isoforms that they use to mediate thrombin signalling, but thrombin is a potent activator of platelets from both species<sup>8</sup>. The observation that ablation of thrombin signalling in mouse platelets yielded protection against thrombosis without spontaneous bleeding, together with the observation that a PAR-blocking antiserum attenuated cyclic flow variations in a monkey model of arterial thrombosis<sup>23</sup>, suggest that inhibition of thrombin signalling in human platelets should be explored as a strategy for the prevention or treatment of thrombosis. □

## Methods

### Inactivation of the *Par4* gene

The pNTK vector<sup>24</sup> was modified by insertion of tandem *loxP* sites<sup>25</sup> 5' and 3' to the *PGK-neo* cassette to generate pNTKloxP. A bacterial artificial chromosome (BAC) from a 129/SvJ mouse genomic library (Genome Systems) was used as a source of DNA containing the mouse *PAR4* gene (*Par4*)<sup>26</sup>. A 9.5-kilobase (kb) long arm that extended from a 5' *NheI* site to a *HindIII* site that was 14 bp 5' to the start codon of *PAR4* was fused to 9 *HindIII/XbaI* fragment containing the  $\beta$ -galactosidase gene. The resulting 12.5-kb fragment was subcloned into pNTKloxP (Fig. 1a). A 1.32-kb short arm was generated from the same BAC by PCR amplification with the primers 5'-CGCTGCTGTATCCTTGGTGC-3' and 5'-TCATGGGACACATAG TAGTAGAT-3'. This fragment, which extended from ten nucleotides 3' of the *PAR4* start codon into exon 2, was inserted between the *PGK-neo* and *PGK-thymidine kinase* (*TK*) cassettes (Fig. 1a) to generate the final targeting construct.

RF8 embryonic stem cells<sup>27</sup> (129/SvJae) were electroporated with the linearized targeting construct. Clones resistant to G418 and FIAU were selected<sup>24</sup> and screened by PCR and Southern blot. PCR primers were 5'-GCTATACGAAGTATCAATTCGCT-3' to the *neo-loxP* junction and 5'-CGCGTACCTTCTCCCTGAATC-3' to *Par4* exon 2 outside of (3' to) short arm (Fig. 1a). Southern analysis used a probe that was PCR amplified from a region of exon 2 that was 3' to that used for the short arm of the targeting vector (Fig. 1a). Two highly chimaeric male mice derived from a *Par4*<sup>+/−</sup> embryonic stem cell clone were bred to C57BL/6 females to generate 13 F<sub>1</sub> *Par4*<sup>+/−</sup> mice. F<sub>2</sub> progeny of F<sub>1</sub> crosses were used for most of the experiments reported.

For studies examining LacZ expression and for some platelet experiments, *Par4*<sup>+/−</sup> mice were crossed with mice bearing a  $\beta$ -actin-Cre recombinase<sup>28</sup> to excise the *neo* cassette. Excision was confirmed by PCR and Southern analysis. Excision of the *neo* cassette did not alter the platelet phenotype.

### Analysis of PAR3 and PAR4 expression

Three micrograms of poly(A)<sup>+</sup> RNA from wild-type and *Par4*<sup>+/−</sup> mouse spleens was analysed by northern blot using a 400-bp *PAR4* probe derived from the 3' end of exon 2 (see above), a 600-bp *EcoRI/SalI* *PAR3* exon 2 probe or a  $\beta$ -actin probe (Clontech) under high stringency. Presence of *PAR4* mRNA in wild-type spleen and its absence in *Par4*<sup>+/−</sup> spleen was also documented by RT-PCR using primers to exon 1 (5'-GAAAGC TTCTGATCTGGCAGCAT-3') and exon 2 (5'-ACTTCGGCTCCTTGAGTT CTACTG-3') (ProStar first strand RT-PCR kit, Stratagene). These tests confirmed ablation of *PAR4* expression in *Par4*<sup>+/−</sup> mice as did functional tests (Fig. 2).

To assess surface expression of *PAR3* in *Par4*<sup>+/−</sup> platelets, mouse platelets were resuspended in platelet buffer (20 mM Tris-HCl buffer at pH 7.4, 140 mM NaCl, 2.5 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mg ml<sup>−1</sup> glucose, 0.5% BSA, 1  $\mu$ M prostaglandin PGE<sub>1</sub> and 5 mM EDTA) and fixed for 10 min with 1% paraformaldehyde in the same. Platelets were then

washed, incubated with 10  $\mu$ g ml<sup>−1</sup> of anti-PAR3 immunoglobulin- $\gamma$  (IgG)<sup>29</sup> at room temperature for 1 h, washed, and incubated with goat anti-rabbit IgG conjugated with fluorescein isothiocyanate (FITC) (Molecular Probes) at 4  $\mu$ g ml<sup>−1</sup> for 30 min. Platelets were then washed twice and analysed by flow cytometry. *Par3*<sup>+/−</sup> platelets were used as a negative control.

### Platelet aggregation and secretion

Anticoagulated blood from three or four *Par4*<sup>+/−</sup> mice (or their wild-type littermates) was pooled for each platelet study. In the experiments shown, washed platelets were prepared and assays performed as previously described<sup>7</sup>. Resuspended platelets were allowed to recover from the action of PGE<sub>1</sub> for at least 20 min before stimulation with thrombin or other agonists. In other studies that examined DIP-thrombin signalling, washed platelets were prepared exactly as described<sup>16</sup>. These studies confirmed the presence of responses to DIP-thrombin in wild-type platelets and the absence of such responses in *Par4*<sup>+/−</sup> platelets.

### Platelet calcium mobilization

Thrombin- and U46619-triggered increases in intracellular Ca<sup>2+</sup> concentration were measured fluorometrically in Fura-2-labelled washed platelets from *Par4*<sup>+/−</sup> and *Par4*<sup>+/−</sup> mice as described previously<sup>29</sup>.

### $\beta$ -galactosidase staining

At 13.5 embryonic days, *Par4*<sup>+/−</sup> (*neo*<sup>−</sup>) or wild-type embryos were fixed for 2 h at 4 °C with 2% paraformaldehyde, 0.5% glutaraldehyde in phosphate-buffered saline, then washed and stained overnight for  $\beta$ -galactosidase<sup>30</sup>. Stained embryos were post-fixed in 2% paraformaldehyde overnight then embedded in plastic. Sections of 8–10  $\mu$ m were examined. Adult spleen and bone marrow were processed similarly and also revealed specific staining in megakaryocytes.

### Tail bleeding times

Bleeding times were performed blind to genotype in mice 3.5–4.5 weeks old. Mice were anaesthetized with pentobarbital. Tails were amputated 5 mm from the tip such that tail artery and veins were all transected. The cut end was immediately immersed in normal saline at 37 °C in a 15-ml clear conical tube. Bleeding was followed visually and time to cessation was recorded. A quantitative estimate of the amount of bleeding was determined by measuring the haemoglobin content of blood collected in saline. Tubes were centrifuged to collect erythrocytes, resuspended in 6 ml of lysis buffer (NH<sub>4</sub>Cl 8.3 g l<sup>−1</sup>; KHCO<sub>3</sub> 1.0 g l<sup>−1</sup>; EDTA 0.037 g l<sup>−1</sup>), and the absorbance of the sample was measured at wavelength 575 nm. The amputated segment of tail was saved for DNA preparation for Southern analysis.

### Mesenteric arteriolar thrombosis

Ferric chloride-triggered thrombosis was assessed essentially as described<sup>17</sup>. Briefly, 3–4-week-old (10–15-g) offspring of *PAR4* heterozygote matings were anaesthetized, the mesentery externalized, and a mesenteric arteriole 70–120  $\mu$ m in diameter was viewed. A piece of Whatman filter paper 1 × 5 mm soaked in saturated ferric chloride solution was placed across the arteriole at time *t* = 0. Blood flow downstream of the injury site was followed by videomicroscopy, and time to the first cessation of flow that lasted >30 s was recorded. Arterioles occluded for 30 s rarely reopened. Mice were genotyped after the thrombosis assays (which were done blind to genotype).

Received 30 May; accepted 25 June 2001.

- Lewis, H. D. Jr *et al.* Protective effects of aspirin against acute myocardial infarction and death in men with unstable angina. Results of a Veterans Administration Cooperative Study. *N. Engl. J. Med.* **309**, 396–403 (1983).
- Coller, B. S., Folts, J. D., Scudder, L. E. & Smith, S. R. Antithrombotic effect of a monoclonal antibody to the platelet glycoprotein IIb/IIIa receptor in an experimental animal model. *Blood* **68**, 783–786 (1986).
- Davies, M. J., Bland, J. M., Hangartner, J. R., Angelini, A. & Thomas, A. C. Factors influencing the presence or absence of acute coronary artery thrombi in sudden ischaemic death. *Eur. Heart J.* **10**, 203–208 (1989).
- CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* **348**, 1329–1339 (1996).
- PURSUIT Trial Investigators (Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy). Inhibition of platelet glycoprotein IIb/IIIa with eptifibatide in patients with acute coronary syndromes. *N. Engl. J. Med.* **339**, 436–443 (1998).
- Davey, M. G. & Lüscher, E. F. Actions of thrombin and other coagulant and proteolytic enzymes on blood platelets. *Nature* **216**, 857–858 (1967).
- Coleman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J. In *Hemostasis and Thrombosis* (eds Coleman, R. W., Marder, V. J., Salzman, E. W. & Hirsh, J.) 3–17 (Lippincott, Philadelphia, 1994).
- Coughlin, S. R. Thrombin signalling and protease-activated receptors. *Nature* **407**, 258–264 (2000).
- Kahn, M. L. *et al.* A dual thrombin receptor system for platelet activation. *Nature* **394**, 690–694 (1998).
- Nakanishi-Matsui, M. *et al.* PAR3 is a cofactor for PAR4 activation by thrombin. *Nature* **404**, 609–613 (2000).
- Kehrel, B. *et al.* Glycoprotein VI is a major collagen receptor for platelet activation: it recognizes the platelet-activating quaternary structure of collagen, whereas CD36, glycoprotein IIb/IIIa, and von Willebrand factor do not. *Blood* **91**, 491–499 (1998).
- Ishihara, H. *et al.* Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* **386**, 502–506 (1997).
- Andrews, R. K. *et al.* The glycoprotein Ib-IX-V complex in platelet adhesion and signaling. *Thromb. Haemost.* **82**, 357–364 (1999).

14. Ware, J. *et al.* Nonsense mutation in the glycoprotein Ib alpha coding sequence associated with Bernard-Soulier syndrome. *Proc. Natl Acad. Sci. USA* **87**, 2026–2030 (1990).
15. De Marco, L., Mazzucato, M., Masotti, A., Fenton, J. W. II & Ruggeri, Z. M. Function of glycoprotein Ib alpha in platelet activation induced by alpha-thrombin. *J. Biol. Chem.* **266**, 23776–23783 (1991).
16. Ramakrishnan, V. *et al.* A thrombin receptor function for platelet glycoprotein Ib-IX unmasked by cleavage of glycoprotein V. *Proc. Natl Acad. Sci. USA* **98**, 1823–1828 (2001).
17. Denis, C. *et al.* A mouse model of severe von Willebrand disease: defects in hemostasis and thrombosis. *Proc. Natl Acad. Sci. USA* **95**, 9524–9529 (1998).
18. Giesen, P. L. *et al.* Blood-borne tissue factor: another view of thrombosis. *Proc. Natl Acad. Sci. USA* **96**, 2311–2315 (1999).
19. Mann, K. G. Biochemistry and physiology of blood coagulation. *Thromb. Haemost.* **82**, 165–174 (1999).
20. Offermanns, S. The role of heterotrimeric G proteins in platelet activation. *Biol. Chem.* **381**, 389–396 (2000).
21. Watson, S., Berlanga, O., Best, D. & Frampton, J. Update on collagen receptor interactions in platelets: is the two-state model still valid? *Platelets* **11**, 252–258 (2000).
22. Schwartz, M. A. & Shattil, S. J. Signaling networks linking integrins and Rho family GTPases. *Trends Biochem. Sci.* **25**, 388–391 (2000).
23. Cook, J. J. *et al.* An antibody against the exosite of the cloned thrombin receptor inhibits experimental arterial thrombosis in the African green monkey. *Circulation* **91**, 2961–2971 (1995).
24. Mortensen, R. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) (Suppl. 23) 9.16.1 (Wiley, New York, 1993).
25. Sauer, B. in *Guide to Techniques in Mouse Development* (eds Wasserman, P. M. & DePamphilis, M. L.) 890–900 (Academic, San Diego, 1993).
26. Kahn, M. L., Hammes, S. R., Botka, C. & Coughlin, S. R. Gene and locus structure and chromosomal localization of the protease-activated receptor gene family. *J. Biol. Chem.* **273**, 23290–23296 (1998).
27. Meiner, V. L. *et al.* Disruption of the acyl-CoA:cholesterol acyltransferase gene in mice: evidence suggesting multiple cholesterol esterification enzymes in mammals. *Proc. Natl Acad. Sci. USA* **93**, 14041–14046 (1996).
28. Lewandoski, M., Meyers, E. N. & Martin, G. R. Analysis of Fgf8 gene function in vertebrate development. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 159–168 (1997).
29. Ishihara, H., Zeng, D., Connolly, A. J., Tam, C. & Coughlin, S. R. Antibodies to protease-activated receptor 3 inhibit activation of mouse platelets by thrombin. *Blood* **91**, 4152–4157 (1998).
30. Schlaeger, T. M., Qin, Y., Fujiwara, Y., Magram, J. & Sato, T. N. Vascular endothelial cell lineage-specific promoter in transgenic mice. *Development* **121**, 1089–1098 (1995).

## Acknowledgements

We thank R. Advincula for mouse husbandry support, I. Hsieh for assistance with transmission electron microscopy of platelets, D. Bainton for examining the electromicrographs, V. Ramakrishnan and D. Phillips for providing advice regarding platelet preparation and assessment of DFP-thrombin signalling, and G. Martin for the  $\beta$ -actin-Cre mouse.

Correspondence and requests for materials should be sent to S.R.C. (e-mail: coughlin@cvrmail.ucsf.edu).

# Mal (MyD88-adaptor-like) is required for Toll-like receptor-4 signal transduction

Katherine A. Fitzgerald<sup>\*†</sup>, Eva M. Palsson-McDermott<sup>\*†</sup>, Andrew G. Bowie<sup>\*‡</sup>, Caroline A. Jefferies<sup>\*</sup>, Ashley S. Mansell<sup>\*</sup>, Gareth Brady<sup>\*</sup>, Elizabeth Brint<sup>\*</sup>, Aisling Dunne<sup>\*</sup>, Pearl Gray<sup>\*</sup>, Mary T. Harte<sup>\*</sup>, Diane McMurray<sup>§</sup>, Dirk E. Smith<sup>§</sup>, John E. Sims<sup>§</sup>, Timothy A. Bird<sup>§</sup> & Luke A. J. O'Neill<sup>\*</sup>

<sup>\*</sup> Department of Biochemistry, Trinity College, Dublin 2, Ireland

<sup>§</sup> Immunex Corporation, Seattle, Washington 98101, USA

<sup>†</sup> These authors contributed equally to this work

The recognition of microbial pathogens by the innate immune system involves Toll-like receptors (TLRs), which recognize pathogen-associated molecular patterns<sup>1–9</sup>. Different TLRs recognize different pathogen-associated molecular patterns, with TLR-4 mediating the response to lipopolysaccharide from Gram-negative bacteria<sup>5–7</sup>. All TLRs have a Toll/IL-1 receptor (TIR) domain,

which is responsible for signal transduction<sup>1,2</sup>. MyD88 is one such protein that contains a TIR domain<sup>10,11</sup>. It acts as an adaptor, being involved in TLR-2, TLR-4 and TLR-9 signalling<sup>12–15</sup>; however, our understanding of how TLR-4 signals is incomplete<sup>15,16</sup>. Here we describe a protein, Mal (MyD88-adaptor-like), which joins MyD88 as a cytoplasmic TIR-domain-containing protein in the human genome. Mal activates NF- $\kappa$ B, Jun amino-terminal kinase and extracellular signal-regulated kinase-1 and -2. Mal can form homodimers and can also form heterodimers with MyD88. Activation of NF- $\kappa$ B by Mal requires IRAK-2, but not IRAK, whereas MyD88 requires both IRAKs. Mal associates with IRAK-2 by means of its TIR domain. A dominant negative form of Mal inhibits NF- $\kappa$ B, which is activated by TLR-4 or lipopolysaccharide, but it does not inhibit NF- $\kappa$ B activation by IL-1RI or IL-18R. Mal associates with TLR-4. Mal is therefore an adaptor in TLR-4 signal transduction.

We were interested in finding an additional adaptor to MyD88 because in cells from MyD88-deficient mice, activation of the transcription factor NF- $\kappa$ B and Jun N-terminal kinase (JNK), although delayed, still occurs in response to lipopolysaccharide (LPS)<sup>16</sup>, as does maturation of dendritic cells<sup>15</sup>. This implies that an additional adaptor is required for a subset of LPS-inducible, TLR-4-dependent signals.

During high-throughput sequencing of a human dendritic cell expressed sequence tag (EST), complementary DNA library we identified a clone (TT4440) that showed sequence similarity to MyD88 and other TIR-domain-containing genes. The nucleotide sequence is available from GenBank (accession number AF406652). By using the radiation hybrid technique, Mal was mapped to chromosome 11q23–24. A murine Mal was also detected as an EST (EMBL accession number AW907851). A full-length open reading frame that encoded Mal was amplified from TT440 template cDNA by direct polymerase chain reaction (PCR) and cloned into the expression vector pDC304, a variant of pDC302 (ref. 17). The antisense primer that we used encoded the haemagglutinin (HA) tag sequence (YPYDVPDYA) immediately before the termination codon.

A PSI BLAST search with Mal identified the known MyD88 proteins from *Homo sapiens*, *Xenopus laevis* and *Mus musculus*, and also TIR-domain-containing proteins including TLR-2 and TLR-4 (Fig. 1a). Similar to MyD88, the carboxy-terminal region of Mal contains a TIR domain. Of note, some core conserved residues in the TIR domain box 1 ((F/Y)D) and box 2 (RD and PG) are present in Mal, whereas box 3 (FW) is absent (Fig. 1a, residues shown in bold). Predictions of the secondary structure of Mal, carried out using the PHD secondary structure prediction program, are consistent with the crystal structure of the TIR domain of TLR-2 (ref. 18). Furthermore, the amino acids in the TIR domain of TLR-2—which form ion pairs (Arg BB3 with Asp BB4 and Glu  $\alpha$ A13) and stabilize the so-called BB loop, which is essential for signalling<sup>18</sup>—are conserved in Mal (Fig. 1a, underlined amino acids). An important difference with MyD88, however, is that the N-terminal region is 75 amino acids shorter and lacks a death domain, which implies that its role in signalling will be different. Using a full-length Mal probe we detected a messenger RNA species of 2.3 kilobases (kb) in most tissues, including peripheral blood leukocytes, spleen, thymus, and most particularly in placenta, liver, kidney, skeletal muscle and heart (Fig. 1b). Identical results were obtained when a probe comprising the TIR domain alone was used (data not shown). This profile was similar to that observed with a full-length MyD88 probe, where a 2.4-kb species was detected. MyD88 mRNA was strongly detected in peripheral blood leukocytes, liver, kidney and spleen.

We also looked for Mal mRNA expression in several cell types by carrying out a PCR with reverse transcription analysis (Fig. 1c). Mal transcripts were detected in mature and immature murine bone-marrow-derived dendritic cells, in the monocytic/macrophage cell lines RAW264, J774 and THP-1, and in Chinese hamster ovary

<sup>‡</sup> Present address: Department of Pharmacology, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland.



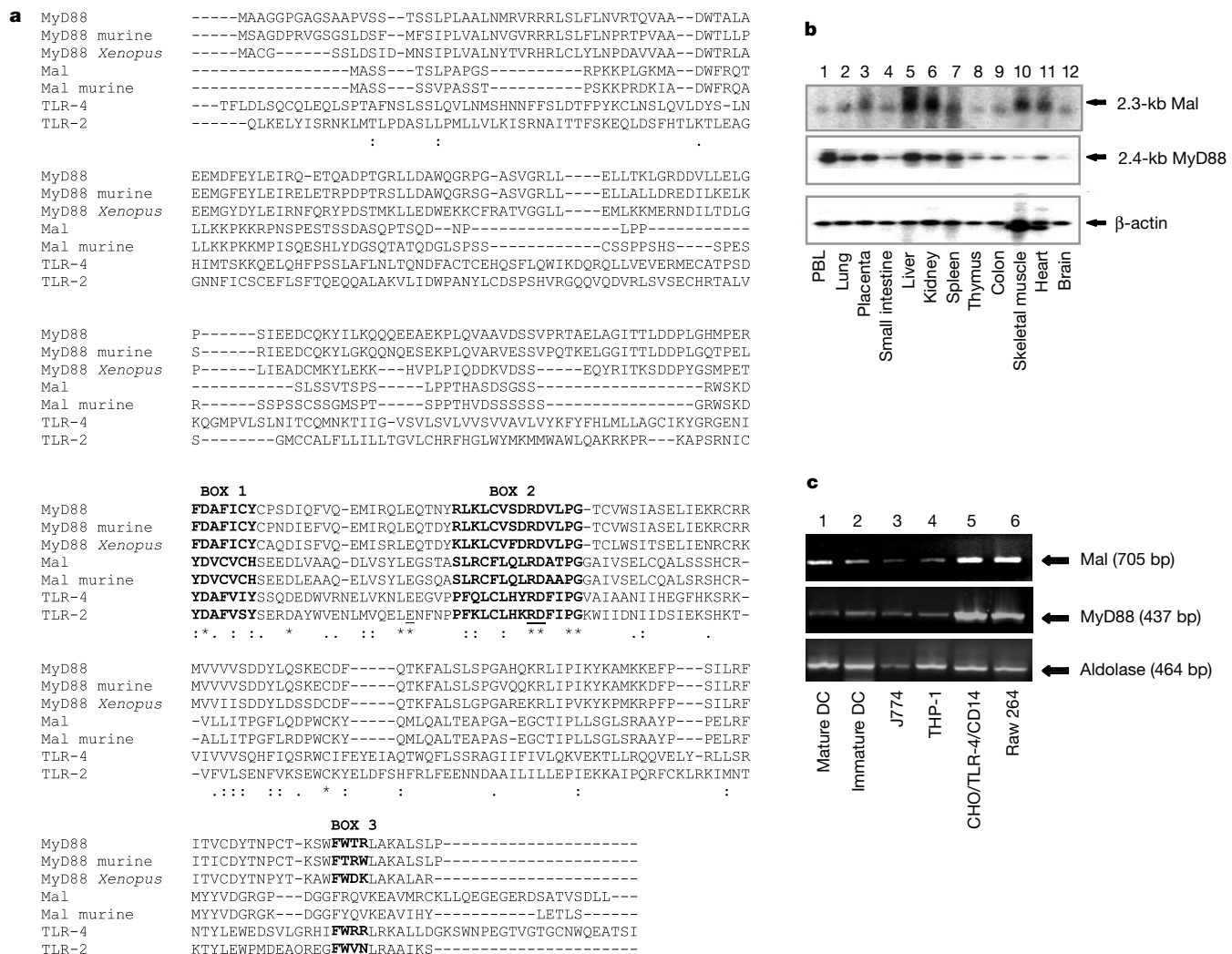
(CHO) cells that were stably transfected with TLR-4 and CD14. MyD88 was also detected in each of these cell lines.

We next carried out functional comparisons between Mal and MyD88. In a similar fashion to overexpression of MyD88, Mal overexpression in human embryonic kidney (HEK) 293 cells activates NF- $\kappa$ B, JNK and the extracellular signal-regulated kinases ERK1 and ERK2 (Fig. 2a–c, respectively). Each effect was dependent on the dose of plasmid that was used (although for NF- $\kappa$ B the effect was less marked at the highest concentration of Mal-encoding plasmid). Each assay was specific for the read outs being monitored. Neither Mal nor MyD88 was able to drive another response, p53 induction, in a similar assay (Fig. 2d). Furthermore we tested the effect of a mutant form of Mal in which Pro 125 in box 2 of the TIR domain is mutated to histidine (Pro125His). This mutation is analogous to that in TLR-4 in C3H/HeJ mice, which is responsible for LPS hyporesponsiveness<sup>7</sup>. Overexpression of this mutant failed to activate NF- $\kappa$ B (not shown), further confirming the specificity of this effect.

The effect of Mal on NF- $\kappa$ B was inhibited by overexpression of the isolated TIR domain from MyD88 (Fig. 3a). The stimulatory effects of MyD88 itself were not inhibited by overexpression of its

own isolated TIR domain<sup>10</sup>. By using the yeast two-hybrid system we found that both Mal and MyD88 were capable of homotypic and heterotypic association (Fig. 3b). We also demonstrated by carrying out co-immunoprecipitations that Mal could associate with MyD88. In extracts of cells that were transfected with epitope-tagged expression constructs of Mal and MyD88—either as full-length proteins (not shown) or consisting only of isolated TIR domains (Fig. 3c, lane 3)—we were able to detect complexes containing both of the proteins by specific immunoprecipitation. The weaker intensity of the band in lane 3 compared with lane 1 possibly suggests that Mal–MyD88 heterodimers are relatively rare compared with Mal homodimers. Furthermore, glutathione S-transferase fusion proteins containing either MyD88 or Mal could isolate Mal from a lysate prepared from cells that overexpressed Mal (not shown).

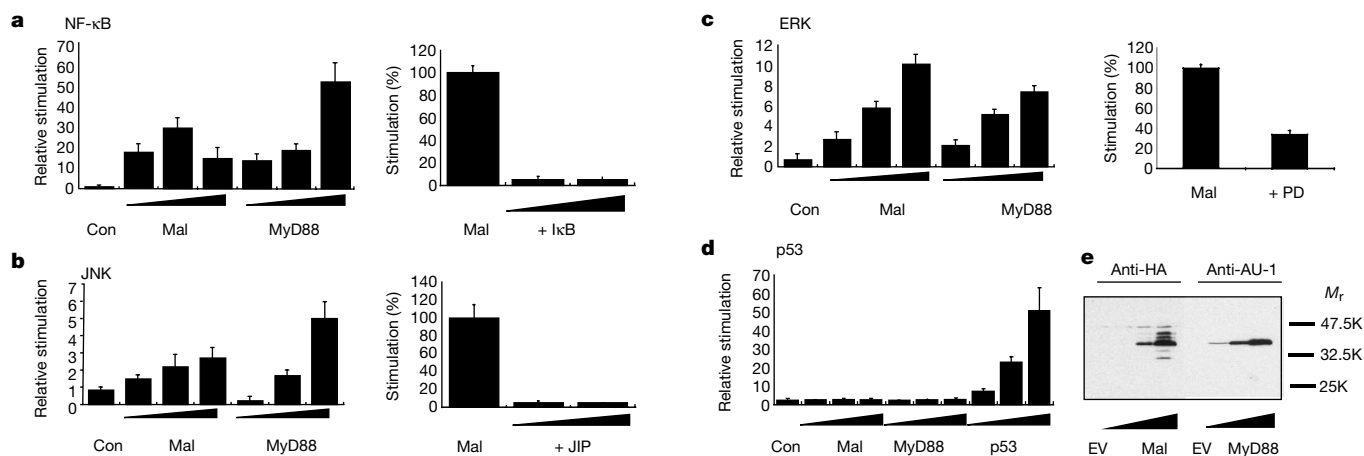
Figure 3d shows the inhibition by dominant negative interleukin (IL)-1 receptor-associated kinase (IRAK) (IRAK 1–211)<sup>19</sup> of NF- $\kappa$ B activation that is induced by MyD88—as shown previously<sup>10,11</sup>—but intriguingly not Mal-induced activation. Co-immunoprecipitation analysis demonstrated that Mal could not interact with IRAK (not shown). MyD88, however, could be co-immunoprecipitated



**Figure 1** MyD88-adaptor-like (Mal) is a widely expressed homologue of MyD88.

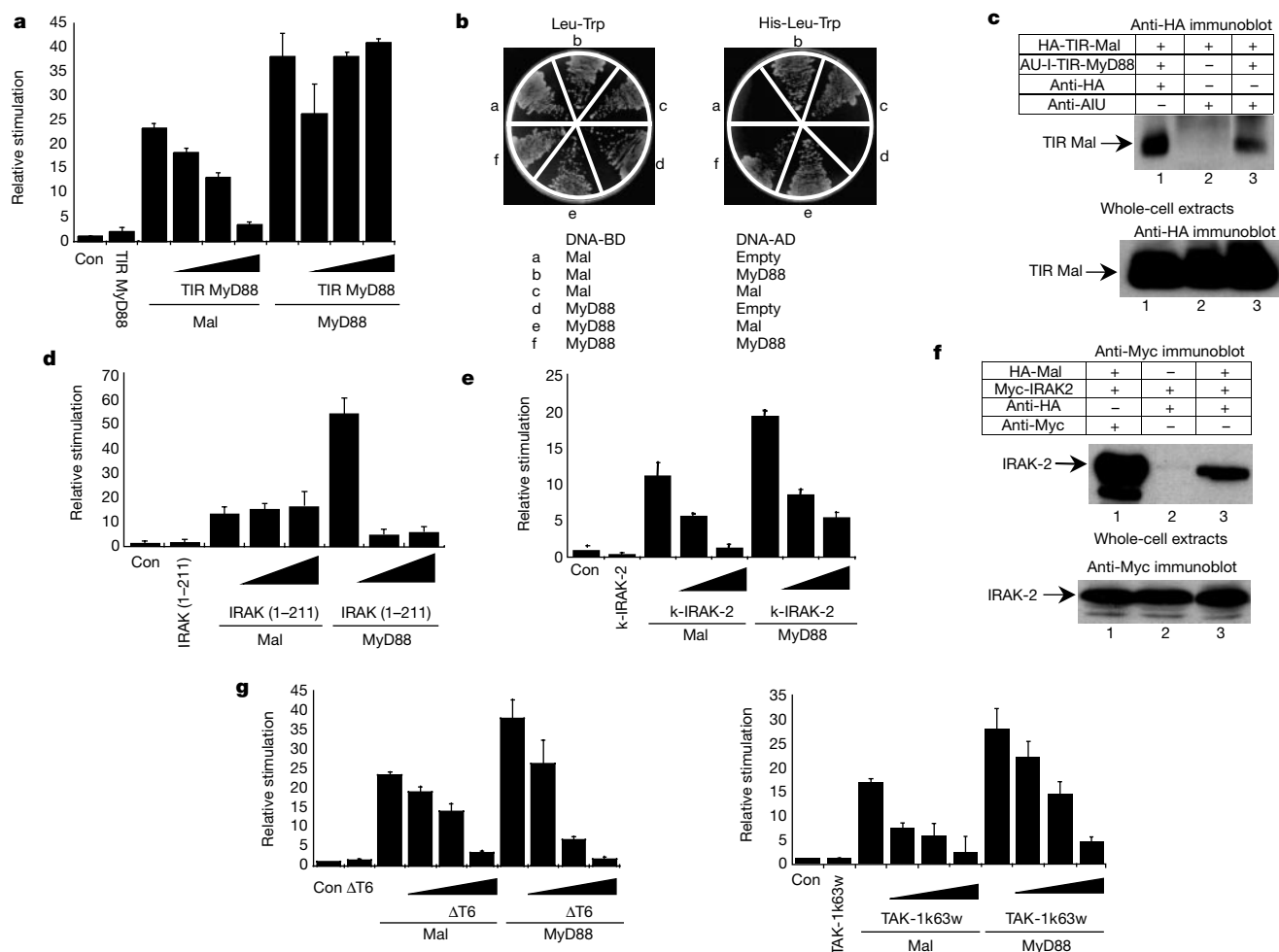
**a**, Alignment of Mal with known MyD88 proteins, TLR-2 and TLR-4. Three regions, box 1, 2 and 3, conserved across all TIR domains and thought to be important in signalling are indicated in bold. Arg BB3, Asp BB4 and Glu  $\alpha$ 13 are all underlined and are likely to form ion pairs, thought essential for signal transduction (see text for details)<sup>18</sup>. Asterisks indicate identical residues in all sequences in the alignment, double dots indicate conserved substitutions, and single dots indicate semi-conserved substitutions.

**b**, Detection of Mal and MyD88 in human adult tissues. Multiple-tissue northern blots (Clontech) were hybridized under stringent conditions with a <sup>32</sup>P-labelled probe corresponding to full-length Mal, full-length MyD88 and  $\beta$ -actin. PBL, peripheral blood leukocytes. **c**, RT-PCR analysis for Mal and MyD88 mRNA. Total RNA was isolated from the indicated cell lines and RT-PCR analysis performed using primers for Mal, MyD88 and aldolase. DC, dendritic cells.



**Figure 2** Mal overexpression activated NF- $\kappa$ B, JNK and ERK1/2. **a–d**, Increasing amounts of Mal or MyD88 expression vectors or p53 (**d**) (1, 40 and 80 ng) were transfected into HEK293 cells. Con, control. Reporter systems were a plasmid harbouring five NF- $\kappa$ B sites upstream of luciferase for NF- $\kappa$ B (**a**); plasmids encoding Gal4 fused to c-Jun (for JNK) (**b**) or Elk1 (for ERK1/2) (**c**) each along with a Gal4 luciferase reporter plasmid; or p53-linked luciferase (**d**). Twenty-four hours after transfection, luciferase reporter gene activity was assayed. Data are expressed as mean relative stimulation  $\pm$  s.d.

s.d. for a representative experiment from four separate experiments, each performed in triplicate. The specificity of each assay was determined by addition of a specific inhibitor for each signal, namely by transfection with plasmids encoding the I- $\kappa$ B super-repressor for NF- $\kappa$ B, or JNK-interacting protein (JIP) for JNK, or by treatment with the MEK inhibitor PD98059 (PD) for ERK1/2. The effect of each inhibitor was specific for each assay (not shown). **e**, Sample western blot demonstrating dose-dependent expression of HA-Mal or AU-1-MyD88. EV, empty vector.



**Figure 3** Mal signals by means of MyD88, IRAK-2, TRAF-6 and TAK. HEK293 cells were transfected with a 5 $\times$  NF- $\kappa$ B reporter gene plasmid and co-transfected with plasmids encoding Mal (20 ng) or MyD88 (20 ng) (**a**, **d**, **e** and **g**), and increasing concentrations (20–80 ng) of the isolated TIR domain from MyD88 (**a**), IRAK (1–211) (**d**), k-IRAK-2 (**e**), TRAF-6 dominant negative ( $\Delta$ T6) or TAK-1 dominant negative (TAK-1k63w) (**g**), for 24 h. Con, control. Mean relative stimulation of luciferase activity  $\pm$  s.d. for a

representative experiment from four separate experiments, each performed in triplicate, is shown. **b**, Yeast two-hybrid analysis of Mal and MyD88 interactions. Leu-Trp plate indicates the presence of either protein, and the His-Leu-Trp plate indicates protein–protein interactions. For immunoprecipitations, HEK293 cells were transfected with 4  $\mu$ g plasmid encoding HA-TIR-Mal or AU-1-TIR-MyD88 (**c**), or encoding HA-Mal or IRAK-2 (**f**). Mal interacts with MyD88 (**c**, lane 3) and IRAK-2 (**f**, lane 3).

with IRAK, where it interacted with the faster-migrating unphosphorylated form of IRAK (not shown; see ref. 11).

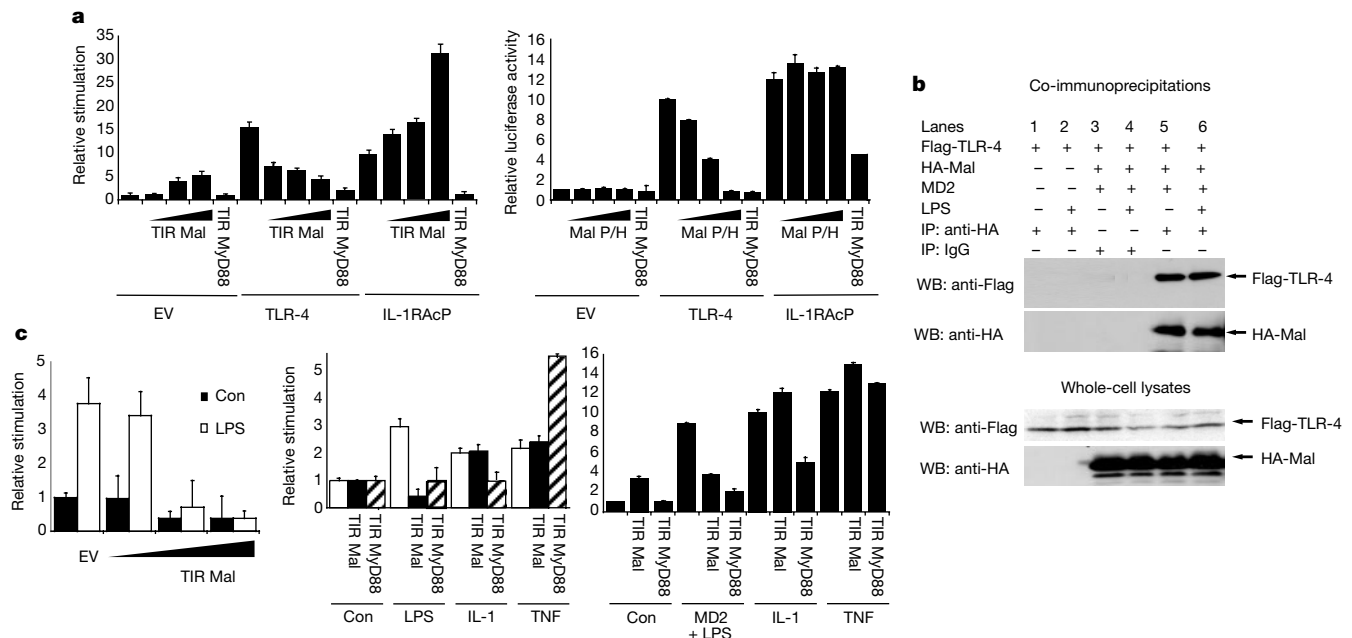
Dominant negative IRAK-2 (k-IRAK-2)<sup>10</sup> blocked the effect of both Mal and MyD88 (Fig. 3e). Figure 3f shows the association between Mal and IRAK-2. This association was observed when either full-length Mal (Fig. 3f) or the isolated TIR domain from Mal (not shown) were immunoprecipitated and the lysates subsequently immunoblotted for IRAK-2. This was also observed when IRAK-2 was immunoprecipitated and probed subsequently for Mal (not shown). MyD88 could also be co-immunoprecipitated with IRAK-2 (ref. 10, and data not shown). These results imply that Mal interacts selectively with IRAK-2. Overexpression of dominant negative forms of both TRAF-6 (ref. 20) (but not TRAF-2 (not shown)) and the protein kinase TAK-1 (ref. 21) inhibited NF- $\kappa$ B activation by Mal or MyD88 in a dose-dependent manner (Fig. 3g). This suggests a role for TRAF-6 and TAK-1 in the signalling pathways transduced by Mal and MyD88.

As residual TLR-4-dependent, LPS-mediated activation of NF- $\kappa$ B was evident in MyD88-deficient mice<sup>16</sup>, we addressed whether Mal might function as an additional adapter for TLR-4. We therefore tested what effect the isolated TIR domain from Mal had on NF- $\kappa$ B activation, which, by analogy with MyD88, should act as a dominant inhibitor. Figure 4a (left panel) demonstrates how increasing concentrations of the isolated TIR domain from Mal activated NF- $\kappa$ B. This response, however, was at no time as strong as that observed with full-length Mal. Overexpression of the isolated TIR domain from MyD88 had no effect; overexpression of TLR-4 in HEK293 cells activates NF- $\kappa$ B. Notably, coexpression with the isolated TIR domain from Mal reduced this signal to the level of this domain alone. The effect was specific for TLR-4 as the response in cells transfected with the IL-1 receptor accessory

protein (IL-1RAcP, as shown), IL-18R or both the type I IL-1 receptor (IL-1RI) and IL-1RAcP (data not shown) were not inhibited. In fact, at the highest concentration tested, the isolated TIR domain of Mal potentiated the response to IL-1RAcP. This was also the case for IL-1RI and IL-1RAcP (not shown). In contrast, each of these was inhibited by the isolated TIR domain from MyD88, as shown for IL-1RAcP in Fig. 4a. We also tested the effect of the mutant form of Mal in which Pro 125 in box 2 of the TIR domain is mutated to histidine (Fig. 4a, right panel). A similar mutation converts TLR-4 into a dominant negative inhibitor<sup>7</sup>. Expression of this mutant does not activate NF- $\kappa$ B. We found, however, that the mutant form of Mal also acted as a dominant negative, blocking the effect of TLR-4, but again not IL-1RAcP, on NF- $\kappa$ B activation. The mutant form of Mal had no effect on NF- $\kappa$ B activation by the receptor-interacting protein (RIP), which is a key participant in tumour-necrosis factor (TNF) signalling (not shown).

We next tested for an interaction between Mal and TLR-4. As shown in Fig. 4b, in HEK293 cells that were stably transfected with Flag-tagged TLR-4 and CD14, HA-Mal could be co-immunoprecipitated with Flag-TLR-4 (lane 5). Treatment of the cells with LPS did not increase the association between Mal and TLR-4, suggesting that the proteins were constitutively associated (compare lane 6 to lane 5). This was the case whether MD2 was also transfected into the cells (as shown in Fig. 4b) or not (not shown). We also demonstrated the reciprocal interaction, whereby Flag-TLR-4 could immunoprecipitate HA-Mal (not shown).

Finally we tested the effect that transfection of the isolated TIR domain from Mal had on NF- $\kappa$ B activation by LPS, in three different cell lines (Fig. 4c). This response was inhibited in a dose-dependent manner in the macrophage cell line THP-1 by the TIR domain of Mal (Fig. 4c, left panel). As we found this cell line to



**Figure 4** Mal is required for TLR-4 and LPS signalling. **a**, HEK293 cells were transfected with a 5 $\times$  NF- $\kappa$ B reporter gene plasmid and co-transfected with plasmids encoding empty vector (EV), TLR-4 (75 ng) or IL-1RAcP (75 ng), and co-transfected with the isolated TIR domain from Mal (10, 40 and 80 ng), MyD88 (80 ng) or Mal Pro125His (Mal P/H) (10, 40 and 80 ng). Mean relative stimulation of luciferase activity  $\pm$  s.d. for a representative experiment from 3–5 separate experiments, each performed in triplicate, is shown.

**b**, Flag-TLR-4/CD14-HEK293 cells transfected with 4  $\mu$ g empty vector (lanes 1 and 2) or plasmid encoding HA-Mal (lanes 3–6), or 2  $\mu$ g of plasmid encoding MD2 (lanes 3–6) were treated with media alone (lanes 1, 3 and 5) or with LPS (200 ng ml<sup>-1</sup>) (lanes 2, 4 and 6) for 15 min. Lysates were immunoprecipitated (IP) with either an anti-HA antibody (lanes 1–2, 5–6) or IgG (lanes 3–4). Mal interacts with TLR-4 (lanes 5–6). Western blotting

(WB) of lysates demonstrates expression of TLR-4 or transfected HA-Mal. **c**, THP-1 (left), RAW264 (middle) or HEK293 cells stably transfected with Flag-TLR-4 (right) were transfected with a 5 $\times$  NF- $\kappa$ B reporter and a plasmid encoding the isolated TIR domain of Mal (10, 50 and 100 ng in THP-1, 100 ng in RAW264 and HEK293-TLR-4), for 24 h. RAW264 or HEK293-TLR-4 cells were also transfected with a plasmid encoding the isolated TIR domain from MyD88 (100 ng), for 24 h. Cells were then left untreated or incubated with LPS (100 ng ml<sup>-1</sup>) from *Escherichia coli* serotype O26:B6 (Sigma), IL-1 (10 ng ml<sup>-1</sup>) or TNF (10 ng ml<sup>-1</sup>) for 6 h. In HEK293-TLR-4 cells, a plasmid encoding MD2 was also transfected as indicated. Mean relative stimulation of luciferase activity  $\pm$  s.d. for a representative experiment from three separate experiments, each performed in triplicate, is shown.



be poorly responsive to IL-1 or TNF, we also tested this effect in the macrophage cell line, RAW264 (Fig. 4c, middle panel) to ensure the specificity of the effect. Again, transfection of the cells with the TIR domain of Mal blocked the LPS response. In these cells, IL-1 and TNF activated NF- $\kappa$ B. Transfection with the TIR domain of Mal had no inhibitory effect on either stimulus. However, the TIR domain from MyD88 inhibited LPS and IL-1, but not TNF, which we found to be potentiated in its effect. Finally, we tested HEK293 cells that were stably transfected with TLR-4 and CD14, which can be rendered responsive to LPS by transfection with MD2 (Fig. 4c, right panel). The cells did not respond to LPS unless they were co-transfected with MD2 (not shown). In these cells, the effects of LPS were again blocked by transfection with the TIR domain of Mal and again, no inhibitory effect was observed against IL-1 or TNF. The TIR domain from MyD88 again blocked LPS and IL-1, but not TNF.

These results support the hypothesis that Mal is a specific adapter for TLR-4 signalling, and provide an important component in the signalling machinery required for the host response to LPS. Signalling by TLR-4 probably involves the adapters MyD88 and Mal, which form dimers, with Mal selectively recruiting IRAK-2 to the complex. Mal may be the adapter responsible for the delayed signalling observed in MyD88-deficient mice<sup>16</sup>. Mal and MyD88 together may be required for a rapid and optimal response. IRAK cannot be activated in MyD88-deficient mice<sup>16</sup>; this is consistent with our data indicating that Mal only interacts with and requires IRAK-2 for signalling, whereas MyD88 requires both IRAKs.

Mal does not participate in IL-1 or IL-18 signalling, and is therefore the first signalling protein found to our knowledge that is not involved in signalling by these other TIR-domain-containing receptors while being required for TLR-4 signal transduction. The lack of involvement of Mal in IL-1 and IL-18 signalling is consistent with a complete unresponsiveness of cells from MyD88-deficient mice to IL-1 and IL-18 (ref. 22). Whether Mal is involved in signalling by other TIR-domain-containing proteins is currently under investigation. Further studies into the function of Mal should provide useful information in the area of LPS signalling, dendritic cell maturation and innate immunity. □

## Methods

### Transfection assays

HEK293 cells ( $2 \times 10^5$  cells per well) were seeded into 96-well plates and transfected as described<sup>23</sup> on the following day with pDC304 Mal-HA or pCDNA3.1 MyD88-AU-1 (1–80 ng), TLR-4 (75 ng) or IL-IRAcP (75 ng), in combination with luciferase reporter genes (80 ng well<sup>-1</sup>  $\kappa$ B-luciferase, p53 luciferase (Stratagene) or a combination of pFR-luciferase (80 ng) and c-jun (2 ng) or Elk1 (5 ng) Gal4 fusion vectors for analysis of JNK or ERK1/2, respectively (Stratagene)). We used the *Renilla*-luciferase reporter gene (40 ng) as an internal control. Where indicated dominant negative expression vectors for MyD88 (MyD88-TIR), Mal (Mal-TIR or Mal Pro125His), IRAK (IRAK 1–211), IRAK-2 (k-IRAK-2), TRAF-6 (T6d) or TAK-1 (TAK-1k63w (TAK-1Lys63Trp))(10, 40 and 80 ng) were co-transfected into the cells.

THP-1 cells ( $1 \times 10^5$  cells per well) were seeded into 24-well plates for 24 h and transfected with FuGENE (Roche) according to the manufacturer's recommendation with an NF $\kappa$ B-luciferase reporter gene (five NF- $\kappa$ B sites upstream of luciferase) (20 ng) and the isolated TIR domain of Mal (10, 50 or 100 ng). RAW264 cells ( $1.5 \times 10^4$  cells per well) were seeded in 96-well plates for 24 h and transfected with Superfect (Qiagen) according to the manufacturers recommendations with 100 ng of an NF $\kappa$ B-luciferase reporter gene and the isolated TIR domain of Mal (100 ng) or MyD88 (100 ng). In all cases, total DNA concentration was kept constant by supplementation with empty vector control. After 24 h, transfected THP-1 cells were either left untreated or stimulated with LPS (200 ng ml<sup>-1</sup>) for 6 h, while transfected Raw264 cells were treated with LPS, IL-1 $\alpha$  or TNF- $\alpha$  (10 ng ml<sup>-1</sup>).

HEK293 cells that stably expressed TLR-4 and CD14 ( $2 \times 10^4$  cells per well) were seeded into 96-well plates and transfected on the following day as described above with MD2 (100 ng) (to render cells responsive to LPS) or empty vector (to analyse IL-1 or TNF responsiveness) in combination with 80 ng well<sup>-1</sup>  $\kappa$ B-luciferase and dominant negative expression vectors for Mal or MyD88.

We used the *Renilla*-luciferase reporter gene (40 ng) as an internal control. In all cases for measurement of luciferase activity, cell lysates were prepared after 24-h transfection and reporter gene activity was measured. Data are normalized for transfection efficiency by dividing firefly luciferase activity with that of *Renilla* luciferase. Data are expressed as mean relative stimulation  $\pm$  s.d. for a representative experiment from a minimum of three separate experiments, each performed in triplicate.

### Transfection and co-immunoprecipitation assays

We transiently transfected HEK293 cells or 293T cells using the FuGENE method with indicated plasmids<sup>23</sup>. The total amount of DNA was kept constant. Twenty-four hours after transfection, cells were lysed in 0.8 ml lysis buffer (50 mM HPES, pH 7.5, 100 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP40 and protease inhibitors). The indicated antibodies were pre-coupled to protein G sepharose, and were incubated with the cell lysates for 2 h. The immune complexes were precipitated and washed thoroughly. The protein was eluted by adding sample buffer and was subsequently fractionated by SDS–polyacrylamide gel electrophoresis (PAGE) and visualized by immunoblotting using the indicated antibodies.

For assessment of TLR-4 and Mal interactions, HEK293 cells that stably expressed Flag-tagged TLR-4 and CD14 were transiently transfected using the FuGENE method with or without MD2 (2  $\mu$ g) or HA-Mal (4  $\mu$ g)<sup>23</sup>. Twenty-four hours after transfection, cells were stimulated with LPS for 15 min or medium alone and then lysed in 1 ml lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.1% NP40, 0.05% CHAPS, 1 mM PMSF and protease inhibitor cocktail (1/100) (Sigma)). HA-Mal immune complexes were precipitated using anti-HA precoupled to protein G sepharose. For control reactions, we used rabbit IgG1. Immunoprecipitated proteins were eluted by adding sample buffer and subsequently fractionated by SDS–PAGE and visualised by immunoblotting using anti-Flag antibodies. Lysates were also immunoblotted for transfected HA-Mal or stably expressed Flag-TLR-4 to demonstrate expression. Anti-Flag-TLR4 immune complexes were precipitated using anti-Flag M2 beads (Sigma), and Flag-TLR4 immune complexes blotted for HA-Mal to demonstrate reciprocal interactions (not shown).

### Northern blotting

The multiple-tissue northern blot was purchased from Clontech and hybridized according to the manufacturer's recommendations with a <sup>32</sup>P-labelled probe corresponding to full-length Mal or full-length MyD88 probes that were generated by restriction enzyme digestion from pDC304 Mal or pCDNA3.1 MyD88 expression vectors. Expression of  $\beta$ -actin was investigated under identical conditions.

### RT-PCR analysis

Total RNA was isolated from amplified granulocyte–macrophage colony-stimulating factor immature or LPS-matured ( $1 \mu$ g ml<sup>-1</sup> for 24 h) murine bone-marrow-derived dendritic cells, RAW264, J774 macrophages or CHO cells that stably expressed TLR-4 and CD14. Total RNA was isolated using TRI reagent (Sigma) according to the manufacturer's recommendations. RT-PCR was then performed using the Enhanced Avian RT-PCR kit (Sigma) with 2  $\mu$ g total RNA as template. The PCR was performed using an annealing temperature of 55 °C with the following primers: (Mal forward, 5'-ATGTCATGGGAATGGCATCATCGACC-3' and reverse 5'-AGACGAGGATCCCTAAAGTATGATCAGA-3', which generates a product of 705 bp; MyD88 TIR forward, 5'-ATAGAATTCGGATCCTAGACCCCTGGGGCAT-3' and reverse 5'-TGTAATC-GAGTC AGGGCAGGGACAA-3', which generates a product of 437 bp; and the house-keeping gene aldolase, aldolase forward 5'-TCCACGACACTGTACCAAGG-3' and reverse 5'-ACCATGTTGGGCTTCAGCAATG-3', which generates a product of 464 bp.

### Yeast two-hybrid analysis

Full-length Mal and MyD88 were sub-cloned into either pGBKT7 (Clontech) downstream of the Gal4 DNA-binding domain (DNA-BD), or into pACT2 (Clontech) downstream of the Gal4 activation domain (DNA-AD). The combinations tested on the Leu-Trip plate indicates the presence of both plasmids, whereas the His-Leu-Trip plate indicates protein–protein interactions.

Received 31 May; accepted 1 August 2001.

- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
- Brightbill, H. D. et al. Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. *Science* **285**, 732–736 (1999).
- Aliprantis, A. O. et al. Cell activation and apoptosis by bacterial lipoproteins through toll-like receptor-2. *Science* **285**, 736–739 (1999).
- Underhill, D. M. et al. The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
- Ozinsky, A. et al. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc. Natl Acad. Sci. USA* **97**, 13766–13771 (2000).
- Poltorak, A. et al. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* **282**, 2085–2088 (1998).
- Hayashi, F. et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
- Hemmi, H. et al. A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–745 (2000).
- Muzio, M., Ni, J., Feng, P. & Dixit, V. M. IRAK (Pelle) family member IRAK-2 and MyD88 as proximal mediators of IL-1 signaling. *Science* **278**, 1612–1615 (1997).
- Wesche, H., Henzel, W. J., Shillinglaw, W., Li, S. & Cao, Z. MyD88: an adapter that recruits IRAK to the IL-1 receptor complex. *Immunity* **7**, 837–847 (1997).
- Kirschning, C. J., Wesche, H., Merrill Ayres, T. & Rothe, M. Human toll-like receptor 2 confers responsiveness to bacterial lipopolysaccharide. *J. Exp. Med.* **188**, 2091–2097 (1998).
- Muzio, M., Natoli, G., Saccani, S., Leviero, M. & Mantovani, A. The human toll signaling pathway: divergence of nuclear factor  $\kappa$ B and JNK/SAPK activation upstream of tumor necrosis factor receptor-associated factor 6 (TRAF6). *J. Exp. Med.* **187**, 2097–2101 (1998).

14. Schnare, M., Holt, A. C., Takeda, K., Akira, S. & Medzhitov, R. Recognition of CpG DNA is mediated by signaling pathways dependent on the adaptor protein MyD88. *Curr. Biol.* **10**, 1139–1142 (2000).
15. Kaisho, T., Takeuchi, O., Kawai, T., Hoshino, K. & Akira, S. Endotoxin-induced maturation of myD88-deficient dendritic cells. *J. Immunol.* **166**, 5688–5694 (2001).
16. Kawai, T., Adachi, O., Ogawa, T., Takeda, K. & Akira, S. Unresponsiveness of MyD88-deficient mice to endotoxin. *Immunity* **11**, 115–122 (1999).
17. Cosman, D. *et al.* Cloning, sequence and expression of human interleukin-2 receptor. *Nature* **312**, 768–771 (1984).
18. Xu, Y. *et al.* Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* **408**, 111–115 (2000).
19. Jefferies, C. *et al.* Transactivation by the p65 subunit of NF- $\kappa$ B in response to interleukin-1 (IL-1) involves MyD88, IL-1 receptor-associated kinase 1, TRAF-6, and Rac1. *Mol. Cell. Biol.* **21**, 4544–4552 (2001).
20. Schwandner, R., Yamaguchi, K. & Cao, Z. Requirement of tumor necrosis factor receptor-associated factor (TRAF)6 in interleukin 17 signal transduction. *J. Exp. Med.* **191**, 1233–1240 (2000).
21. Yamaguchi, K. *et al.* Identification of a member of the MAPKKK family as a potential mediator of TGF- $\beta$  signal transduction. *Science* **270**, 2008–2011 (1995).
22. Adachi, O. *et al.* Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143–150 (1998).
23. Bowie, A. *et al.* A46R and A52R from vaccinia virus are antagonists of host IL-1 and toll-like receptor signaling. *Proc. Natl Acad. Sci. USA* **97**, 10162–10167 (2000).

# Acknowledgements

We thank K. Mills for DC samples; M. Muzio for plasmids encoding MyD88, IRAK-2 and TLR-4; E. L. Cooke for plasmids encoding IRAK (1–211); W. Falk for plasmids encoding IL-1R1 and IL-1RAcP; and H. Sakurai at Tanabe Seiyaku, Japan, for plasmids encoding TAK-1k63w. We also thank D. Caffrey for advice on alignments, E. Kurt-Jones for HEK293 cells stably transfected with Flag-TLR-4 and CD14, and P. Henneke and D. Golenbock for helpful discussions. This work was supported by the European Union Biomed, Biotech and TMR programmes, Enterprise Ireland and the National Pharmaceutical Biotechnology Centre.

Correspondence and requests for materials should be addressed to L.A.J.O. (e-mail: laoneill@tcd.ie).

# Loss of *p16<sup>Ink4a</sup>* confers susceptibility to metastatic melanoma in mice

Paul Krimpenfort<sup>†</sup>, Kim C. Quon<sup>†</sup>, Wolter J. Mooi<sup>‡</sup>, Ate Loonstra<sup>\*</sup> & Anton Berns<sup>\*</sup>

<sup>\*</sup> Division of Molecular Genetics and Centre for Biomedical Genetics, and <sup>‡</sup> Department of Pathology, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

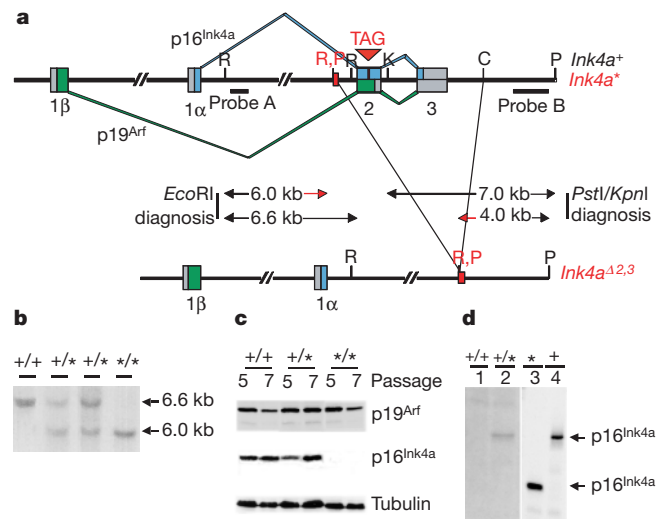
<sup>†</sup> These authors contributed equally to this work

*CDKN2A* (*INK4a/ARF*) is frequently disrupted in various types of human cancer, and germline mutations of this locus can confer susceptibility to melanoma and other tumours<sup>1</sup>. However, because *CDKN2A* encodes two distinct cell cycle inhibitory proteins, *p16<sup>Ink4a</sup>* and *p14<sup>ARF</sup>* (*p19<sup>Arf</sup>* in mice)<sup>2</sup>, the mechanism of tumour suppression by *CDKN2A* has remained controversial. Genetic disruption of *Cdkn2a* (*p19<sup>Arf</sup>*) (hereafter *Arf*) alone predisposes mice to tumorigenesis<sup>3</sup>, demonstrating that *Arf* is a tumour-suppressor gene in mice. We mutated mice specifically in *Cdkn2a* (*p16<sup>Ink4a</sup>*) (hereafter *Ink4a*). Here we demonstrate that these mice, designated *Ink4a<sup>Δ2,3</sup>*, do not show a significant predisposition to spontaneous tumour formation within 17 months. Embryo fibroblasts derived from them proliferate normally, are mortal, and are not transformed by oncogenic *HRAS*. The very mild phenotype of the *Ink4a<sup>Δ2,3</sup>* mice implies that the very strong phenotypes of the original *Ink4a/Arf<sup>Δ2,3</sup>* mice were primarily or solely due to loss of *Arf*. However, *Ink4a<sup>Δ2,3</sup>* mice that are deficient for *Ink4a* and heterozygous for *Arf* spontaneously develop a wide spectrum of tumours, including melanoma. Treatment of these mice with the carcinogen 7,12-dimethylbenzanthracene (DMBA) results in an increased incidence of melanoma, with

frequent metastases. Our results show that, in the mouse, *Ink4a* is a tumour-suppressor gene that, when lost, can recapitulate the tumour predisposition seen in humans.

The unrelated proteins *p16<sup>Ink4a</sup>* and *p14<sup>ARF</sup>* are generated by alternative first-exon usage and alternative reading frames<sup>2</sup> (see Fig. 1a). *p16<sup>Ink4a</sup>* arrests cells in the G1 phase of the cell cycle by binding the cyclin-dependent kinases CDK4 and CDK6 and inhibiting their ability to phosphorylate and inactivate the retinoblastoma (RB1) family of tumour-suppressor proteins<sup>1,4</sup>. *p14<sup>ARF</sup>* acts through HDM2 (Mdm2 in mice) to stabilize and activate the key checkpoint protein TRP53, which can arrest cells in both G1 and G2 or induce apoptosis<sup>2</sup>. Although most mutations at the *CDKN2A* locus found in human tumours result in disruption of both the *ARF* and *INK4A* transcripts, the presence of point mutations and small deletions that specifically affect *p16<sup>Ink4a</sup>* strongly implicate *p16<sup>Ink4a</sup>* as a tumour suppressor<sup>1,5</sup>. To functionally test this hypothesis, we generated mice carrying a point mutation specifically affecting *p16<sup>Ink4a</sup>*. This allele, designated *Ink4a<sup>\*</sup>*, is silent in the *p19<sup>Arf</sup>* reading frame, but introduces a stop codon in *p16<sup>Ink4a</sup>* at conserved amino-acid position 101 (Fig. 1a), resulting in deletion of the fourth ankyrin repeat motif. The analogous human allele (*CDKN2A-W110X*) is a naturally occurring mutation found in a wide variety of human tumour types<sup>1</sup>, and results in an unstable protein that is severely defective in its ability to inhibit phosphorylation of RB1 and to induce cell-cycle arrest in transfected cells<sup>6</sup>.

*Ink4a<sup>Δ2,3</sup>* mice are born in the expected mendelian ratios, are fertile, and do not show any gross morphological or behavioural abnormalities (data not shown). The presence of the *Ink4a<sup>\*</sup>* allele in these mice was confirmed by Southern blotting (Fig. 1b). Immunoblotting and immunoprecipitation experiments showed that wild-type *p16<sup>Ink4a</sup>* protein was absent in *Ink4a<sup>Δ2,3</sup>* primary mouse embryo fibroblasts (MEFs) (Fig. 1c, d). Levels of *p19<sup>Arf</sup>* were unaffected. No truncated *p16<sup>Ink4a</sup>* protein product could be detected, either by immunoblotting (not shown) or by immunoprecipitation of radio-labelled cell extracts (Fig. 1d), suggesting that, like its human analogue, the *p16<sup>Ink4a</sup>* mutant protein is unstable. Failure to detect *p16<sup>Ink4a</sup>* was not due to the inability of the antibody to recognize the truncated protein, as the same antibody readily



**Figure 1** Generation of *Ink4a* mutant mice. **a**, The *Ink4a/Arf* locus and *Ink4a<sup>\*</sup>* allele (top) and the exon 2–3 deletion in the *Ink4a<sup>Δ2,3</sup>* allele (bottom). Changes in the mutant alleles are shown in red; the red boxes are the *loxP* sites. C, *Csp451*; K, *KpnI*; P, *PstI*; R, *EcoRI*; TAG, stop codon in *Ink4a<sup>\*</sup>* allele. **b**, Germline transmission of the *Ink4a<sup>\*</sup>* allele. Southern blot of *EcoRI*-digested tail DNA from offspring of a *Ink4a<sup>Δ2,3</sup>* brother/sister cross using probe A. **c**, Immunoblots of cell lysates from MEFs at the indicated passage numbers. **d**, Immunoprecipitation of <sup>35</sup>S-Met/Cys-labelled MEF cell extracts (lanes 1 and 2) or *p16<sup>Ink4a</sup>* translated *in vitro* (lanes 3 and 4), using antibody against *p16<sup>Ink4a</sup>*.

detected p16<sup>Ink4a</sup> protein translated *in vitro* (Fig. 1d).

The absence of detectable protein suggests that *Ink4a*<sup>\*</sup> is a null or severely hypomorphic allele. To functionally confirm this, we transfected wild-type or mutant *Ink4a* into U2OS osteosarcoma cells. Whereas transfection of the wild-type allele caused a 32.6% ± 1.1 (mean ± s.e.) increase in the percentage of cells in the G1 phase of the cell cycle, *Ink4a*<sup>\*</sup> failed to cause a significant increase (7.7% ± 9.0), demonstrating that *Ink4a*<sup>\*</sup> is defective in inducing cell-cycle arrest in G1.

MEFs mutated in components of the Rb1 or Trp53 pathways are defective in various aspects of proliferative control, displaying altered proliferation rates, cell-cycle distributions and cell size compared with wild-type cells<sup>3,7–11</sup>. Surprisingly, *Ink4a*<sup>+/+</sup> MEFs displayed no major changes in any of these properties (Fig. 2a, b, and not shown). Furthermore, like wild-type cells, *Ink4a*<sup>+/+</sup> MEFs arrested in response to serum starvation and γ-irradiation (data not shown). These results suggest that loss of p16<sup>Ink4a</sup> does not strongly affect cell proliferation or G1 cell-cycle checkpoints under standard growth conditions *in vitro*.

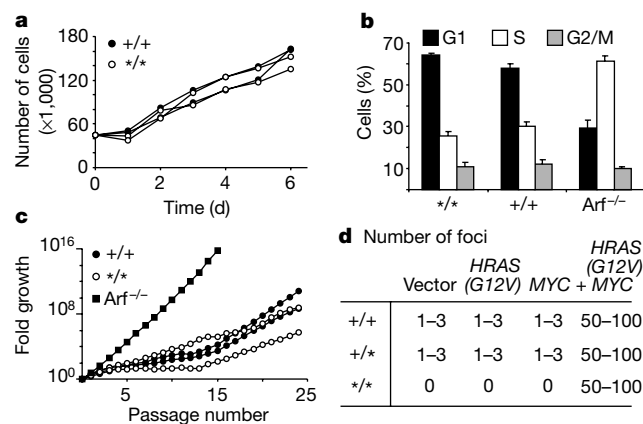
Cellular stresses, such as prolonged *in vitro* culturing or activation of the *HRAS* proto-oncogene, can induce a senescence-like state in mammalian cells<sup>12</sup>. p16<sup>Ink4a</sup> and its downstream targets, the RB1 family members, have been implicated in this response in both human<sup>13,14</sup> and mouse cells<sup>8–11,15</sup>, although in mouse cells the p19<sup>Arf</sup>–Mdm2–Trp53 pathway is also essential<sup>3,7,9</sup>. To directly test the role of p16<sup>Ink4a</sup> in senescence, we passaged MEFs according to a 3T3 immortalization protocol<sup>16</sup>. Under these conditions, *Ink4a*<sup>+/+</sup>, *Ink4a*<sup>+/+</sup> and wild-type MEFs grew indistinguishably from each other, proliferating rapidly on explantation into culture, before entering a senescence period with slow growth from which immortalized cultures ultimately emerged (Fig. 2c, data not shown for heterozygotes). In contrast, control *Arf*<sup>−/−</sup> MEFs did not senesce, and in immortalized cultures of all genotypes, expression of either p19<sup>Arf</sup> or Trp53 was altered in a manner consistent with disruption of the pathway (see Supplementary Information Fig. 1). These results argue that loss of p16<sup>Ink4a</sup> is insufficient for MEF immortalization, and suggest instead that cellular senescence in MEFs is mediated primarily by the p19<sup>Arf</sup>–Mdm2–Trp53 pathway and/or by a Rb1-family-dependent pathway in which p16<sup>Ink4a</sup> is not essential.

The protein p16<sup>Ink4a</sup> has been proposed to prevent transformation by oncogenic *HRAS* (*HRAS*-G12V) in primary cells by inducing a senescence-like response<sup>8,14</sup>. However, retroviral transduction

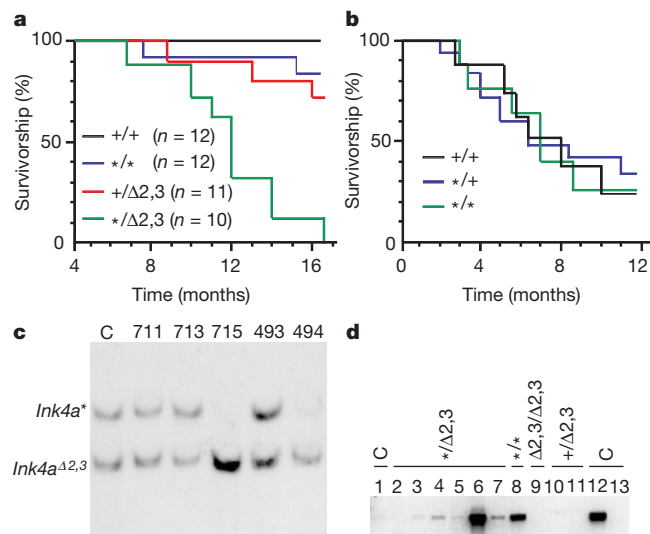
of *HRAS*-G12V into *Ink4a*<sup>+/+</sup> MEFs failed to result in transformed foci (Fig. 2d). Furthermore, mutation of *Ink4a* did not enhance the ability of *HRAS*-G12V and *MYC* together to form foci. These results indicate that loss of *Ink4a* is insufficient to allow transformation by *HRAS*-G12V, and further argue against a major role for p16<sup>Ink4a</sup> in replicative or oncogene-induced senescence in MEFs.

*Ink4a*<sup>+/+</sup> mice show only a subtle predisposition to spontaneous tumour formation that is not yet statistically significant: only 2 out of 12 (17%) *Ink4a*<sup>+/+</sup> mice have developed tumours (B-cell lymphomas) within 17 months (Fig. 3a), whereas age-matched control wild-type mice have remained free of tumours (*P* = 0.15). To determine whether loss of *Ink4a* can collaborate with other oncogenic mutations in tumorigenesis, we crossed the *Ink4a*<sup>\*</sup> allele onto different tumour-predisposing genetic backgrounds. Neither heterozygous nor homozygous *Ink4a*<sup>\*</sup> mutations significantly accelerated B-cell lymphomagenesis induced by Eμ-*Myc* (Fig. 3b), and tumours arising in heterozygous animals failed to show loss of the wild-type *Ink4a* allele (data not shown). This argues against a significant role for *Ink4a* in Eμ-*Myc*-induced lymphomagenesis, consistent with the susceptibility of *Ink4a*<sup>+/+</sup> MEFs to *Myc*-induced apoptosis (K.Q., unpublished data), and previous observations that *Rb1* heterozygosity also fails to significantly collaborate with Eμ-*Myc* in tumorigenesis<sup>17</sup>.

Deletion of one copy of the *CDKN2A* locus and intragenic mutation in the remaining *INK4a* allele is a combination frequently found in human tumours<sup>1</sup>. To model this condition, we generated *Ink4a*<sup>Δ2,3</sup> mice. The *Ink4a*<sup>Δ2,3</sup> allele deletes *Ink4a* exons 2 and 3, thereby disrupting both p16<sup>Ink4a</sup> and p19<sup>Arf</sup> function (see Methods). *Ink4a*<sup>Δ2,3</sup> mice lacking intact *Ink4a* but retaining one functional *Arf* allele showed markedly accelerated tumorigenesis compared with age-matched heterozygote *Ink4a*<sup>+/Δ2,3</sup> controls (Fig. 3a). All *Ink4a*<sup>Δ2,3</sup> mice developed tumours within 17 months (*n* = 10), giving rise to various sarcomas, carcinomas, lymphomas and metastatic melanoma, with an average latency of about 12 months, whereas only 25% of control mice (*n* = 12) developed tumours within this time (*P* < 0.00005; Table 1). Importantly, the *Ink4a*<sup>\*</sup> allele was retained in 3 out of 5 spontaneous tumours



**Figure 2** Growth characteristics of *Ink4a*<sup>+/+</sup> MEFs. **a**, Proliferation curves of *Ink4a*<sup>+/+</sup> and wild-type MEFs. Cultures of two embryos of each genotype are shown. **b**, Cell cycle profiles of MEFs. *Ink4a* genotype is indicated. **c**, Growth curve of MEFs cultured under a modified 3T3 protocol. The number of passages after thawing of frozen MEF stocks is indicated. *Ink4a*<sup>+/+</sup> and wild-type MEFs were originally frozen at passage 3, *Arf*<sup>−/−</sup> MEFs at passage 11. Representative cultures of two embryos of each *Ink4a* genotype are shown. **d**, Number of foci induced by infection with *HRAS*-G12V and/or *MYC*-containing retroviruses.



**Figure 3** *Ink4a* is a tumour-suppressor gene. **a**, Survival curve of untreated mice of the indicated *Ink4a* genotypes. **b**, Survival curve of Eμ-*Myc* transgenic mice in the *Ink4a* background indicated. **c**, Loss-of-heterozygosity analysis of spontaneous tumours arising from *Ink4a*<sup>Δ2,3</sup> mice by Southern blotting using probe B (Fig. 1a). C, control tail DNA. **d**, Expression of *Arf* mRNA in tumours, as detected by RT-PCR. Lanes 2–4: spontaneous tumours 711, 713 and 493, respectively. Lanes 5–10: DMBA-induced tumours. Controls (C): lane 1, normal skin and tail; lane 12, wild-type MEF expressing *HRAS*-G12V; lane 13, no RNA.



**Table 1 Spontaneous tumour development in *Ink4a*<sup>+/Δ2,3</sup> and *Ink4a*<sup>\*/Δ2,3</sup> mice**

| Mouse no. | Sex | <i>Ink4a</i> genotype | Age at death† (months) | Site of tumour        | Tumour type                            |
|-----------|-----|-----------------------|------------------------|-----------------------|----------------------------------------|
| 411       | F   | +/ $\Delta$ 2,3       | 9                      | Skin                  | Squamous cell carcinoma                |
| 426       | F   | +/ $\Delta$ 2,3       | 13                     | Liver                 | Histiocytic sarcoma                    |
| 402       | M   | +/ $\Delta$ 2,3       | 16                     | Lung                  | Adenocarcinoma                         |
| 1590      | M   | */ $\Delta$ 2,3       | 7                      | Skin                  | Fibrosarcoma                           |
| 713       | M   | */ $\Delta$ 2,3       | 10                     | Skin, peritoneal wall | Angiosarcoma                           |
| 721       | M   | */ $\Delta$ 2,3       | 10                     | Liver, skin           | Lymphoma                               |
| 715       | F   | */ $\Delta$ 2,3       | 11                     | Uterus                | Histiocytic sarcoma                    |
| 494       | M   | */ $\Delta$ 2,3       | 12                     | Skin                  | Spindle cell carcinoma                 |
| 493       | M   | */ $\Delta$ 2,3       | 12                     | Spleen, liver         | Metastasizing melanoma                 |
| 400       | F   | */ $\Delta$ 2,3       | 12                     | Liver, spleen, lung   | Histiocytic sarcoma                    |
| 403       | M   | */ $\Delta$ 2,3       | 14                     | Bladder, subcutaneous | Sarcoma, B-cell lymphoma               |
| 396       | M   | */ $\Delta$ 2,3       | 14                     | Lung                  | Adenocarcinoma                         |
| 711       | F   | */ $\Delta$ 2,3       | 17                     | Lung, uterus          | Adenocarcinoma, stromal cell carcinoma |

† See Methods.

originating from *Ink4a*<sup>\*/Δ2,3</sup> mice (Fig. 3c), thus excluding allelic loss of *Arf* as an underlying mechanism in these tumours. No silencing of the *Arf* promoter by methylation could be detected in these tumours (representative examples are shown in Supplementary Information Fig. 2). *Arf* messenger RNA could be detected by polymerase chain reaction with reverse transcription (RT-PCR) in 2 out of the 3 tumours (Fig. 3d, lanes 2–4), as well as in 4 out of 4 DMBA-induced tumours (Fig. 3d, lanes 5–8, described below) retaining a wild-type *Arf* allele (not shown). No *Arf* transcripts were detectable in tumours arising from *Ink4a*<sup>Δ2,3/Δ2,3</sup> animals, nor in *Ink4a*<sup>+/Δ2,3</sup> tumours that had lost the wild-type allele (Fig. 3d,

lanes 9–11, and not shown). These results demonstrate that, in mice, *Ink4a* is a tumour-suppressor gene the loss of which can cooperate with *Arf* heterozygosity in the formation of a wide variety of tumour types.

Melanoma is the predominant tumour type seen in human families carrying germline mutations at the *CDKN2A* locus<sup>1</sup>. To increase the incidence of melanoma in *Ink4a*<sup>\*/Δ2,3</sup> mice, we applied a single dose of DMBA (see Methods). Treatment induced a variety of tumours, including skin papillomas and pulmonary adenocarcinomas. Strikingly, 7 out of 14 animals developed cutaneous melanocytic tumours between 3 and 9 months of age. The smallest lesions strongly resembled human benign blue naevi (Fig. 4a). Larger lesions, which usually extended diffusely into underlying fat and muscle, were far more densely cellular and showed varying nuclear atypia together with increased mitoses (Fig. 4b). Foci of markedly increased cellularity and mitotic activity were encountered (Fig. 4c). In 3 out of 8 melanocytic tumours (2 DMBA-induced and 1 spontaneous tumour), metastases to the lymph nodes, lungs, spleen or liver (Fig. 4d–f) were detected and histologically confirmed as true malignant melanomas. Histologically similar melanomas were observed in DMBA-treated *Ink4a*<sup>\*/+</sup> mice, although limited numbers prevented a detailed comparison of latencies. No melanomas were observed in DMBA-treated control animals, either in those followed by us for 8 months or in historical controls<sup>3,18,19</sup>.

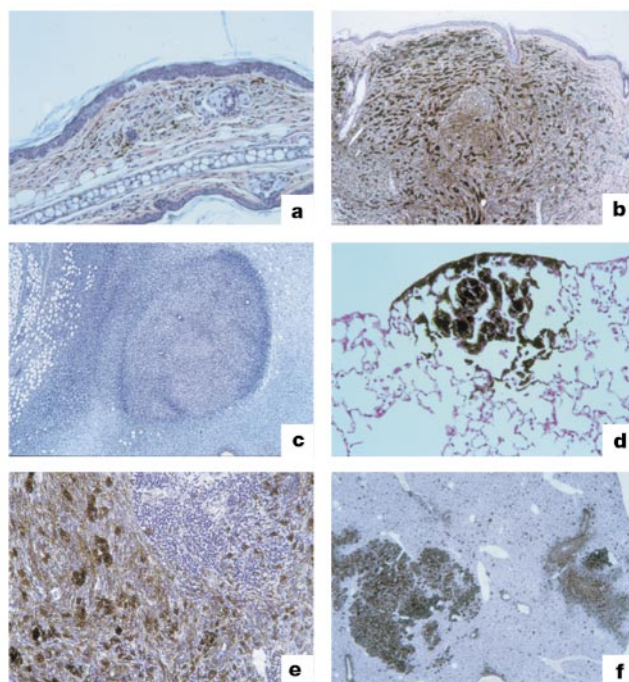
The spontaneous tumour susceptibility of *Ink4a*<sup>\*/Δ2,3</sup> mice relative to *Ink4a*<sup>\*/+</sup> animals indicates that *Arf* heterozygosity contributes to tumorigenesis in *Ink4a*<sup>\*/Δ2,3</sup> mice ( $P = 0.0004$ ). Interestingly, the wild-type *Arf* allele was frequently retained and expressed in tumours arising from these mice (Fig. 3c, d), in contrast to previous reports using mice with functional p16<sup>Ink4a</sup>, in which *Arf* was almost invariably lost<sup>17,19–21</sup>. Thus, in the absence of functional p16<sup>Ink4a</sup>, reduced dosage of the *Arf* gene may be sufficient to contribute to tumorigenesis, possibly reflecting a genetic interaction between *Arf* and the *Ink4a*–*Rb1* pathway, such as that proposed by Carnero *et al.*<sup>9</sup>. Regardless of the underlying molecular mechanism, our results raise the possibility that *ARF* heterozygosity may also contribute to human tumorigenesis<sup>22</sup>, and may account for why mutations specifically affecting the remaining *ARF* allele have only rarely been described in human tumours carrying hemizygous *CDKN2A* deletions<sup>23,24</sup>.

We demonstrate here that although loss of p16<sup>Ink4a</sup> function does not strongly affect cell proliferation, senescence, or *HRAS*-induced transformation in cultured MEFs, *Ink4a* can nevertheless act as a tumour-suppressor gene in an intact mouse. These findings highlight our present lack of understanding of the signals that induce p16<sup>Ink4a</sup> during tumorigenesis *in vivo*, and emphasize the need for appropriate animal models of p16<sup>Ink4a</sup> function. The *Ink4a* mutant mice we describe here represent, to our knowledge, the first example of a mouse model for an inherited cancer syndrome that both mimics the genetic changes underlying the syndrome, and recapitulates the process of tumour progression as seen in humans, from benign lesion to the end stage of metastasis. □

## Methods

### Generation of *Ink4a* mutant mice

A clone carrying mouse *Ink4a* exons 2 and 3 was isolated from a 129/Ola λGEM12 genomic library using exon 2 as a probe. A 110-base mutagenic primer was used to introduce a TGG-to-TAG nonsense mutation at codon 101 in the *Ink4a* reading frame in exon 2. PCR was performed with the mutagenic primer and SP6 primer, using as a template an exon 2 containing a 1.8-kilobase (kb) *KpnI* fragment subcloned from the λGEM12 genomic clone into pSP72 (Promega). The introduced mutation generates a *StyI* restriction site for genotyping. A pGK-neo selection cassette flanked by *loxP* sites ('floxed') was inserted in the unique *BamHI* site 1 kb upstream of exon 2, and the mutated locus was recombined into the original genomic clone, from which a 10-kb *XbaI* fragment was used as a targeting vector to electroporate IB10 (129/Ola) embryonic stem cells<sup>25</sup>. G418-resistant embryonic stem-cell colonies were analysed for homologous recombination by Southern blotting, as in Fig. 1b. *StyI* digestion and sequencing of PCR fragments covering exon 2 confirmed the presence of the introduced nonsense mutation. Two positive clones were injected into B6 blastocysts, and resulting chimaeric males were bred to FVB/N females to



**Figure 4** Histology of melanocytic tumours in *Ink4a*<sup>\*/Δ2,3</sup> mice. **a**, Small blue naevus, consisting of an intradermal focus of slender, elongated, pigmented melanocytes. **b**, Larger, densely cellular melanocytic skin tumour, mitotically active (not evident at this magnification). **c**, Large, partly non-pigmented melanocytic tumour diffusely extending into subcutaneous fat (left), with focus with increased cellularity (centre). Pigmentation is not evident in this field. **d**, Small, heavily pigmented melanoma lung metastasis. **e**, Spontaneous melanoma metastasis in the spleen. **f**, Multiple pigmented liver tumours with histology identical to that of splenic tumour (same animal as in **e**). Distribution of metastases along portal tracts suggests haematogenous spread from the splenic tumour.

achieve germline transmission. The *neo* selection marker was removed by crossing with Deleter Cre mice<sup>26</sup>, and sibling progeny were crossed to generate *Ink4a*<sup>+/+</sup> mice.

The *Ink4a*<sup>Δ2,3</sup> allele was derived by crossing *Ink4a*<sup>lox</sup> mice with Deleter Cre mice. We constructed the *Ink4a*<sup>lox</sup> allele by inserting *loxP* sites 1 kb upstream of *Ink4a* exon 2 and 1.5 kb downstream of exon 3. Embryonic stem cells and breeding schedules used to generate *Ink4a*<sup>lox/lox</sup> mice were as for *Ink4a*<sup>+/+</sup> mice. Homozygous *Ink4a*<sup>Δ2,3</sup> mice and MEFs are phenotypically identical to the published *Ink4a* knockout mice and MEFs<sup>18</sup> (data not shown).

## Biochemical analyses

Genotyping was performed by Southern blotting and/or by exon-2-specific PCR combined with *SbfI* digestion. Probe A is a 500-base pair (bp) fragment at the 5' end of the original AGEN12 genomic clone outside of the *XbaI* fragment used for targeting; probe B is identical to the 1.3-kb *BamHI* probe described by Serrano *et al.*<sup>18</sup>. Antibodies used were p16<sup>INK4a</sup> and M-156 (Santa Cruz); p19<sup>ARF</sup> and R562 (Abcam); and tubulin and YL1/2 (ECACC). For immunoprecipitation assays, cells were metabolically labelled with 400 mCi <sup>35</sup>S-Met/Cys for 90 min in Met- and Cys-free DMEM containing 1% dialysed serum. Proteins translated *in vitro* were generated by the TNT-coupled transcription/translation system (Promega) from linear PCR products containing a wild-type or mutated *Ink4a* complementary DNA. For RT-PCR of *Arf*, 1 μg total RNA was primed with oligo-dT and reverse transcribed with Superscript II (Life Technologies). One-fortieth of the RT reaction was amplified with the following primers and cycling conditions: GAGGGTTTCTTGTTAAAGTTCG (exon 1β forward primer); and GAGGCCGATTAGCTCTGCTC (exon 3 reverse primer); 35 cycles of 94 °C (1 min), 55 °C (45 s) and 72 °C (2 min).

## Cell culture

MEFs were isolated from 13.5-d embryos, and grown at 5% CO<sub>2</sub> in DMEM (Gibco) supplemented with glutamine, penicillin/streptomycin, 10% FCS and 0.1 mM β-mercaptoethanol. All experiments were performed on MEFs before passage 7 unless otherwise indicated. Passage 8 *Arf*<sup>-/-</sup> MEFs<sup>3</sup> were obtained from C. J. Sherr via D. S. Peeper. For 3T3 immortalization protocols<sup>16</sup>, 9 × 10<sup>5</sup> cells were plated in 9-cm plates every 3–4 d. Cells were counted with a Casy-1 cell counter. For cell-cycle profile analyses, cells were labelled for 4 h with 10 μM bromodeoxyuridine (BrdU). BrdU incorporation and DNA content was measured with anti-BrdU antibody conjugated with fluorescein isothiocyanate (FITC, Becton Dickinson) and propidium iodide (5 μg ml<sup>-1</sup>) on a FACScan flow cytometer according to the manufacturer's instructions. For focus formation assays, cells were infected with pBABEpuro retroviral vectors<sup>27</sup> containing human *HRAS-G12V* or *MYC*, or no insert, selected for 3 d on puromycin (2 μg ml<sup>-1</sup>), plated at 1.5 × 10<sup>5</sup> cells per well in six well plates, and observed after 18 d. Doubly infected cells were infected first with *MYC* retrovirus and puromycin selected, and subsequently infected with *HRAS-G12V* virus without selection. Infection efficiencies were >50%. Retroviruses were produced with ΦNX-E cells obtained from G. Nolan. All MEF experiments were performed at least in duplicate using at least two independent MEF cultures from different pregnancies, and compared with a control group containing at least one sibling embryo culture.

## Tumorigenesis studies

Mice were killed when seriously ill or when visible tumours reached a diameter of 1 cm. DMBA treatment was performed as described<sup>18</sup>. Statistical analysis of tumour susceptibility was performed with SPSS 9.0 software, using a log rank test applied to Kaplan–Meier survival curves. Animals that died without visible tumours were removed from the curves (*n* = 3). Tumours were fixed in formalin and sections were stained with haematoxylin and eosin. Samples from which sufficient quantities of relatively pure tumour-cell DNA could be isolated were analysed for loss of heterozygosity.

Received 12 June; accepted 10 July 2001.

- Ruas, M. & Peters, G. The p16<sup>INK4a</sup>/CDKN2A tumor suppressor and its relatives. *Biochim. Biophys. Acta* **1378**, F115–F177 (1998).
- Sherr, C. J. Tumor surveillance via the ARF–p53 pathway. *Genes Dev.* **12**, 2984–2991 (1998).
- Kamijo, T. *et al.* Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19<sup>ARF</sup>. *Cell* **91**, 649–659 (1997).
- Sherr, C. J. & Roberts, J. M. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.* **13**, 1501–1512 (1999).
- Sharpless, N. E. & DePinho, R. A. The INK4A/ARF locus and its two gene products. *Curr. Opin. Genet. Dev.* **9**, 22–30 (1999).
- Arav, P., Knudsen, E. S., Wang, J. Y., Cavenee, W. K. & Huang, H. J. Point mutations can inactivate *in vitro* and *in vivo* activities of p16(INK4a)/CDKN2A in human glioma. *Oncogene* **14**, 603–609 (1997).
- Harvey, M. *et al.* *In vitro* growth characteristics of embryo fibroblasts isolated from p53-deficient mice. *Oncogene* **8**, 2457–2467 (1993).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16<sup>INK4a</sup>. *Cell* **88**, 593–602 (1997).
- Carnero, A., Hudson, J. D., Price, C. M. & Beach, D. H. p16<sup>INK4a</sup> and p19<sup>ARF</sup> act in overlapping pathways in cellular immortalization. *Nature Cell Biol.* **2**, 148–155 (2000).
- Sage, J. *et al.* Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev.* **14**, 3037–3050 (2000).
- Dannenberg, J. H., van Rossum, A., Schuijff, L. & te Riele, H. Ablation of the *retinoblastoma* gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev.* **14**, 3051–3064 (2000).
- Wright, W. E. & Shay, J. W. Cellular senescence as a tumor-protection mechanism: the essential role of counting. *Curr. Opin. Genet. Dev.* **11**, 98–103 (2001).

- Kiyono, T. *et al.* Both Rb/p16<sup>INK4a</sup> inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**, 84–88 (1998).
- Zhu, J., Woods, D., McMahon, M. & Bishop, J. M. Senescence of human fibroblasts induced by oncogenic Raf. *Genes Dev.* **12**, 2997–3007 (1998).
- Peeper, D. S., Dannenberg, J. H., Douma, S., te Riele, H. & Bernards, R. Escape from premature senescence is not sufficient for oncogenic transformation by Ras. *Nature Cell Biol.* **3**, 198–203 (2001).
- Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963).
- Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13**, 2670–2677 (1999).
- Serrano, M. *et al.* Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
- Kamijo, T., Bodner, S., van de Kamp, E., Randle, D. H. & Sherr, C. J. Tumor spectrum in ARF-deficient mice. *Cancer Res.* **59**, 2217–2222 (1999).
- Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. Disruption of the ARF–Mdm2–p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev.* **13**, 2658–2669 (1999).
- Jacobs, J. J. *et al.* Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* **13**, 2678–2690 (1999).
- Rizos, H., Becker, T. M., Holland, E. A., Kefford, R. F. & Mann, G. J. Differential expression of p16<sup>INK4a</sup> and p16β transcripts in B-lymphoblastoid cells from members of hereditary melanoma families without CDKN2A exon mutations. *Oncogene* **15**, 515–523 (1997).
- Randerson-Moor, J. A. *et al.* A germline deletion of p14(ARF) but not CDKN2A in a melanoma-neural system tumour syndrome family. *Hum. Mol. Genet.* **10**, 55–62 (2001).
- Burri, N. *et al.* Methylation silencing and mutations of the p14<sup>ARF</sup> and p16<sup>INK4a</sup> genes in colon cancer. *Lab. Invest.* **81**, 217–229 (2001).
- Robanus-Maandag, E. *et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. *Genes Dev.* **12**, 1599–1609 (1998).
- Schwenk, F., Baron, U. & Rajewsky, K. A *cre*-transgenic mouse strain for the ubiquitous deletion of *loxP*-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
- Morgenstern, J. P. & Land, H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.* **18**, 3587–3596 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

## Acknowledgements

We thank D. Peeper and M. van Lohuizen for advice and retroviral vectors, M. van der Valk for histological analysis, K. Rajewsky for the Deleter Cre mice, C. Sherr for *Arf*<sup>-/-</sup> MEFs, G. Nolan for ΦNX-E retroviral producer cells, and J. Vink and K. van Veen for technical assistance, and the Netherlands Cancer Institute animal facility for providing animal care. Part of this work was supported by the Dutch Cancer Society.

Correspondence and requests for materials should be addressed to A.B. (e-mail: [therns@nki.nl](mailto:therns@nki.nl)).

# Loss of p16<sup>INK4a</sup> with retention of p19<sup>ARF</sup> predisposes mice to tumorigenesis

Norman E. Sharpless\*, Nabeel Bardeesy\*, Kee-Ho Lee\*†, Daniel Carrasco\*‡, Diego H. Castrillon\*‡, Andrew J. Aguirre\*, Emily A. Wu\*, James W. Horner\* & Ronald A. DePinho\*

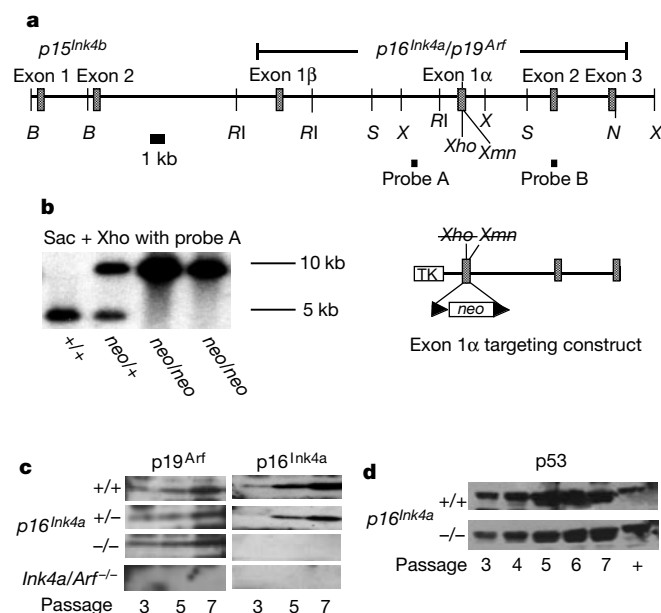
\* Departments of Adult Oncology, Medicine and Genetics, Harvard Medical School and the Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA  
‡ Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

The cyclin-dependent kinase inhibitor p16<sup>INK4a</sup> can induce senescence of human cells, and its loss by deletion, mutation or epigenetic silencing is among the most frequently observed molecular lesions in human cancer<sup>1,2</sup>. Overlapping reading frames in the *INK4A/ARF* gene encode p16<sup>INK4a</sup> and a distinct tumour-suppressor protein, p19<sup>ARF</sup> (ref. 3). Here we describe the generation and characterization of a p16<sup>INK4a</sup>-specific knockout

† Present address: Laboratory of Molecular Oncology, Korea Cancer Center Hospital, Seoul 139-706, Korea.

mouse that retains normal  $p19^{Arf}$  function. Mice lacking  $p16^{Ink4a}$  were born with the expected mendelian distribution and exhibited normal development except for thymic hyperplasia. T cells deficient in  $p16^{Ink4a}$  exhibited enhanced mitogenic responsiveness, consistent with the established role of  $p16^{Ink4a}$  in constraining cellular proliferation. In contrast to mouse embryo fibroblasts (MEFs) deficient in  $p19^{Arf}$  (ref. 4),  $p16^{Ink4a}$ -null MEFs possessed normal growth characteristics and remained susceptible to Ras-induced senescence. Compared with wild-type MEFs,  $p16^{Ink4a}$ -null MEFs exhibited an increased rate of immortalization, although this rate was less than that observed previously for cells null for  $Ink4a/Arf$ ,  $p19^{Arf}$  or  $p53$  (refs 4, 5). Furthermore,  $p16^{Ink4a}$  deficiency was associated with an increased incidence of spontaneous and carcinogen-induced cancers. These data establish that  $p16^{Ink4a}$ , along with  $p19^{Arf}$ , functions as a tumour suppressor in mice.

Standard targeting methods were used to generate a germline exon-1 $\alpha$  deletion that removed the translation start site and the first 35 codons of  $p16^{Ink4a}$  while leaving intact all known *cis*-regulatory and splice elements (Fig. 1a). Mice heterozygous for this mutant allele were crossed with EIIa-Cre mice<sup>6</sup> to excise the *loxP*-flanked neomycin-resistance cassette. We chose to delete the selectable marker given the proximity of  $p15^{Ink4b}$  and  $p19^{Arf}$ , as retained selection cassettes have been shown to alter expression of genes over 100 kilobases (kb) distant<sup>7</sup>. Correct targeting of exon 1 $\alpha$  was confirmed by Southern blotting (probes A and B, Fig. 1b), by multiplex polymerase chain reaction (PCR) specific for null (designated  $p16^{Ink4a}/-$ ), wild-type or neomycin-containing alleles (see Supplementary Information Fig. 1), and by direct sequencing of exon 1 $\alpha$  and 2 in  $p16^{Ink4a}/-$  and  $+/+$  MEFs. Moreover, western-blot analysis of cultured MEFs revealed the elimination or 50% reduction in  $p16^{Ink4a}$  levels in  $p16^{Ink4a}/-$  and  $+/-$  cells, respectively

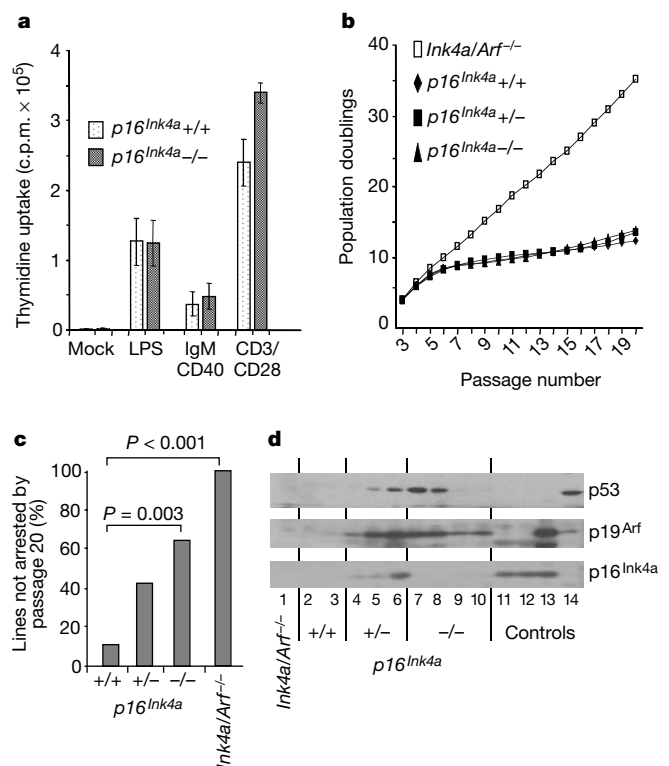


**Figure 1** Deletion of  $p16^{Ink4a}$ . **a**, The murine *Ink4a/Arf/Ink4b* locus and exon-1 $\alpha$  targeting strategy. The neomycin-resistance cassette (*neo*), flanked by *loxP* (solid triangles), was cloned into an *Eco*RI–*Not*I fragment at *Xho*I and *Xmn*I sites in exon 1 $\alpha$ . These sites, which flank the translation start site of  $p16^{Ink4a}$ , were destroyed in the targeted allele. *B*, *Bam*HI; *N*, *Not*I; *R*, *Eco*RI; *S*, *Sac*I; *X*, *Xba*I; TK, thymidine kinase. **b**, Southern analysis of wild-type and *neo* alleles. Successful targeting was determined by *Sac*I/*Xho*I double digestion using probe A, yielding a fragment 5.0-kb longer in the neomycin-containing allele. **c**, Western analysis of  $p16^{Ink4a}$  and  $p19^{Arf}$  in serially passaged MEFs. **d**, Western analysis showing equivalent p53 stabilization in  $p16^{Ink4a}/+$  and  $-/-$  MEFs with serial passaging. +, ultraviolet-irradiated MEFs.

(Fig. 1c, Supplementary Information Fig. 5). Both products of the murine *Ink4a/Arf* locus,  $p16^{Ink4a}$  and  $p19^{Arf}$ , have been shown to accumulate in MEFs as a function of passage in culture<sup>4,8</sup>, and rising  $p19^{Arf}$  levels correlate with stabilization and accumulation of p53. This passage-induced accumulation of  $p19^{Arf}$  and p53 was similar for all three  $p16^{Ink4a}$  genotypes (Fig. 1c). Lastly, anti- $p16^{Ink4a}$  or anti-Cdk4 immunoprecipitates of <sup>35</sup>S-labelled MEF lysates failed to detect  $p16^{Ink4a}$  or an aberrant truncated product in  $p16^{Ink4a}/-$  MEFs (data not shown).

Mice with the genotype  $p16^{Ink4a}/-$  arose from heterozygous intercrosses at the expected frequency (57:127:76,  $+/+$ : $+/-$ : $-/-$ ) and histological surveys of the spleen, liver, lung, kidney, stomach, intestine, testis and brain ( $n = 3$  per genotype, ages 20–30 weeks) did not reveal morphological anomalies. Serial body-mass determinations were normal among 35 littermates (11  $p16^{Ink4a}/-$ ). Complete blood counts and red-blood-cell indices were comparable at 24 weeks of age. Flow cytometric analysis of spleen showed no significant differences in the distribution of lymphoid, myeloid, erythroid or granulocytic lineages (seven matched pairs of  $p16^{Ink4a}/+$  and  $-/-$  mice, ages 8–30 weeks; data not shown).

Thymuses of  $p16^{Ink4a}/-$  mice were larger than those of  $p16^{Ink4a}/+$  littermate controls (57 mg versus 33 mg, age 8–10 weeks,  $n = 3$  matched pairs) with increased cortical thickness (0.34 mm versus 0.24 mm,  $n = 4$  matched pairs; see Supplementary Information Fig. 3) and total cellularity ( $10.7 \times 10^7$  versus  $6.8 \times 10^7$  cells,  $n = 3$  matched pairs). Although the absolute numbers of  $CD4^+CD8^-$ ,  $CD4^-CD8^+$  and  $CD4^+CD8^+$  cells were increased in  $p16^{Ink4a}/-$



**Figure 2** Growth of  $p16^{Ink4a}$ -null cells. **a**, Thymidine uptake (mean  $\pm$  s.e.m.) of total splenocytes after stimulation with the indicated mitogens. c.p.m., counts per minute. **b**, Averaged results of 54 different MEF cultures (18  $+/+$ , 14  $+/-$ , 14  $-/-$  and 8 *Ink4a/Arf<sup>-/-</sup>*) passaged on a 3T9 protocol. For averaging, the growth of senesced lines was considered to be 0 population doublings per passage. **c**, Enhanced growth in culture by reduced  $p16^{Ink4a}$  gene dosage, shown as lines proliferating beyond passage 20 in a 3T9 assay. *P*-values were calculated pairwise by a two-sided Fisher's exact test. **d**, Western blot analysis of immortalized clones. Results from 10 representative lines of indicated genotypes are shown. Controls: lanes 11–12,  $p19^{Arf}$ -null cell lines; lane 13, p53-negative tumour line; lane 14,  $p16^{Ink4a}/-$  H-Ras-transduced cell line.



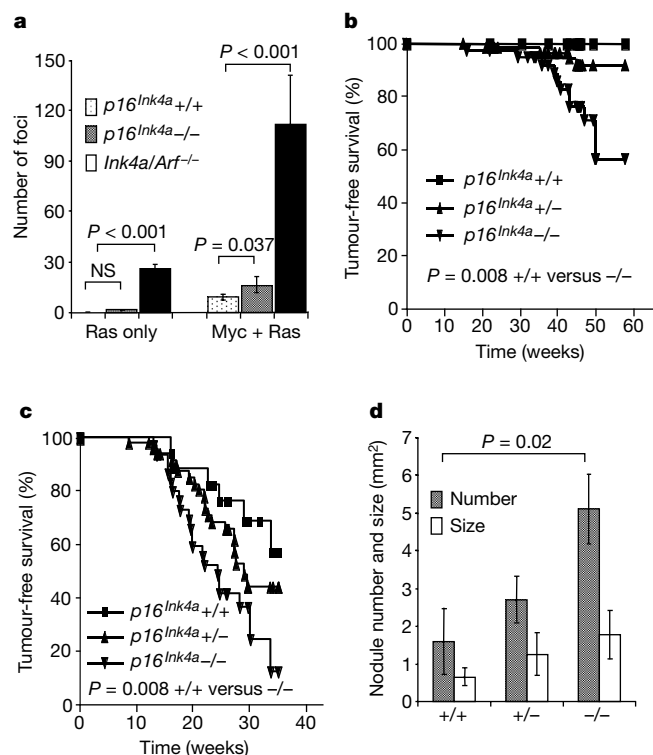
thymuses relative to littermate controls; this hyperplasia was most marked in the double-positive compartment ( $6.4 \times 10^7$  versus  $3.5 \times 10^7$  double-positive cells in  $p16^{Ink4a}/-/-$  and  $+/+$  thymuses, respectively). The increase in double-positive cells was seen in mice aged 8–61 weeks ( $n = 13$  matched pairs). The above findings, coupled with the established role of  $p16^{Ink4a}$  in cellular proliferation, prompted an analysis of the mitogenic responsiveness of lymphocytes *in vitro*. Increased proliferation was observed in unsorted splenocytes after stimulation with CD3 and CD28, activators of mature T cells, whereas no differences were seen with the B-cell mitogens lipopolysaccharide (LPS) and immunoglobulin- $\mu$  (IgM) with CD40 (Fig. 2a,  $n = 4$  matched pairs, aged 8–9 weeks). These results point to a role for  $p16^{Ink4a}$  in thymic development and T-cell proliferation.

In contrast to T lymphocytes, multiple, independently derived MEF cultures at early (<8) passages exhibited indistinguishable growth rates during a 6-day culture period (data not shown). Similarly,  $p16^{Ink4a}$  status did not affect the cell-cycle profiles of unsynchronized, exponentially growing cultures; colony formation following low-density seeding; growth rates in medium with reduced serum content (0.5%); or S-phase re-entry after release from serum starvation (Supplementary Information Figs 3 and 4; and data not shown). Consistent with similar growth rates of early-passage MEFs, Rb-phosphorylation patterns, Cdk4 activity and complex composition were unaffected by  $p16^{Ink4a}$  status (data not shown).

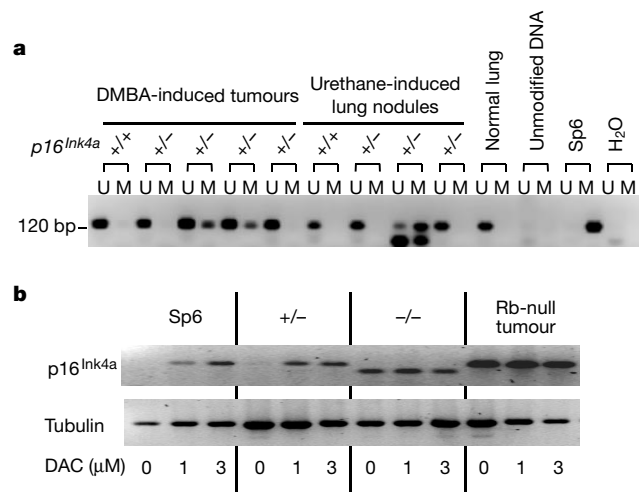
Accumulation of  $p16^{Ink4a}$  correlates with the onset of senescence in human cells<sup>9</sup>, and its loss is required for immortalization of certain

human cell types, even in the setting of p53 inactivation<sup>10,11</sup>. MEFs deficient for both  $p16^{Ink4a}$  and  $p19^{Arf}$  or  $p19^{Arf}$  alone<sup>4</sup> exhibit more rapid growth rates and are more readily immortalized in culture, are susceptible to transformation by activated H-Ras alone, and exhibit resistance to Ras-induced senescence<sup>4,12</sup>. To examine the role of  $p16^{Ink4a}$  in these processes, long-term *in vitro* growth of MEFs was assayed by a 3T9 protocol (Fig. 2b). The proliferative rates of both  $p16^{Ink4a}/-/-$  and  $+/+$  cells decreased with serial passage, whereas  $Ink4a/Arf^{-/-}$  cells showed no culture-induced replicative arrest. Furthermore, senescence-associated  $\beta$ -galactosidase activity, a surrogate marker of senescence<sup>13</sup>, was comparable between  $p16^{Ink4a}/+/+$  and  $-/-$  cells when measured at multiple passages (for example, 28–40% versus 23–33% positive at passage 8). Similarly,  $p16^{Ink4a}$  status did not affect Ras-induced senescence of early-passage MEFs, with all genotypes exhibiting  $p19^{Arf}$  and p53 activation, senescent morphology, and senescence-associated  $\beta$ -galactosidase positivity (Supplementary Information Fig. 5 and data not shown). In contrast, Ras-transduced  $Ink4a/Arf^{-/-}$  MEFs did not undergo growth arrest or stabilize p53, but rather were transformed in this setting (data not shown). These data confirm that  $p19^{Arf}$ , and not  $p16^{Ink4a}$ , is the principal mediator of passage- and Ras-induced growth arrest in murine cells.

Although proliferative rates declined comparably with serial passage in both  $p16^{Ink4a}/-/-$  and  $+/+$  cells (Fig. 2b), immortalized lines of  $p16^{Ink4a}/-/-$  MEFs emerged with greater facility than  $+/+$  littermate cultures when assayed on a 3T9 protocol using a rigorous definition of immortalization (see Methods, Fig. 2c and Supplementary Information Fig. 8). Inactivation of  $p19^{Arf}$  or loss of p53 has been determined previously to be a common means of immortalization of wild-type MEFs<sup>4</sup>, and we analysed the expression of these genes in 26 representative immortalized lines (3  $Ink4a/Arf^{-/-}$ , 7  $p16^{Ink4a}/+/+$ , 7  $+/-$  and 9  $-/-$ ). Evidence of loss of  $p19^{Arf}$  expression or p53 function was seen in 7 out of 7  $p16^{Ink4a}/+/+$  and 5 out of 7  $+/-$  lines, but in only 3 out of 9  $-/-$  lines (Fig. 2d;  $P = 0.01$  for  $+/+$  versus  $-/-$  lines). In 6 out of 9  $p16^{Ink4a}/-/-$  lines,  $p19^{Arf}$  was detectable but p53 expression was undetectable or very low (for example, lanes 9 and 10 in Fig. 2d). This pattern is not consistent with p53 mutation, as  $p19^{Arf}$  is elevated in cells lacking functional



**Figure 3**  $p16^{Ink4a}$ -null mice are tumour prone. **a**, Transformation assays. Averaged foci number after transfection of MEFs with Ras and/or Myc. Results shown are the average ( $\pm$  s.e.m.) of at least two independent experiments, with four lines per genotype.  $P$ -values were calculated with a two-sided  $t$ -test. **b**, Tumour-free survival of untreated littermates (26  $p16^{Ink4a}/+/+$ , 60  $+/-$  and 39  $-/-$ ). Survival curves in **b** and **c** were compared pairwise by the log rank test. **c**, Tumour-free survival of DMBA-treated littermates (18  $p16^{Ink4a}/+/+$ , 48  $+/-$  and 29  $-/-$ ) exposed at 5–7 days old. **d**, Lung nodule number and size (mean  $\pm$  s.e.m.) after urethane treatment of littermate mice (5  $p16^{Ink4a}/+/+$ , 10  $+/-$  and 8  $-/-$ ). Ranges were compared with an unpaired  $t$ -test.



**Figure 4** Methylation of  $p16^{Ink4a}$  in tumours. **a**, Bisulphite-treated tumour DNA was amplified with primers specific for either an unmethylated (U) or methylated (M) allele. The detection of products from both primer sets in some samples is indicative of normal tissue contamination. Normal lung and unmodified DNA are shown as controls; Sp6 is a methylated cell line<sup>24</sup>. bp, base pairs. **b**, Expression of  $p16^{Ink4a}$  in tumour cell lines after treatment with 0, 1 or 3  $\mu$ M DAC, a methylation inhibitor. By PCR with reverse transcription, an increase in  $p16^{Ink4a}$  transcription is detected after treatment with DAC in  $p16^{Ink4a}/+/-$  lines. As the primers span the region deleted in the null allele, a smaller transcript is detected in the  $p16^{Ink4a}/-/-$  lines.

p53 (refs 3, 4). In 2 out of 4 such lines, p53 induction was demonstrated after ultraviolet irradiation (see Supplementary Information Fig. 6) with concomitant Rb hypophosphorylation and growth arrest. Expression of the Rb protein was documented in all immortalized lines, and persistent p16<sup>Ink4a</sup> expression could be seen in all p16<sup>Ink4a</sup>+/- lines (for example, lanes 4–6 in Fig. 2d). The functional integrity of p16<sup>Ink4a</sup> in these lines was confirmed by the presence of p16<sup>Ink4a</sup> in anti-Cdk4 immunoprecipitates (data not shown). Our data suggest that p19<sup>Arf</sup>, and not p16<sup>Ink4a</sup>, mediates the passage-induced entry of MEFs into a phase of slow growth that is variably termed 'senescence'. Although not sufficient for immortalization, loss of p16<sup>Ink4a</sup> does seem to facilitate the escape of this growth arrest, occasionally without direct functional compromise of p19<sup>Arf</sup> and p53. These data are reminiscent of results from several other groups<sup>14–17</sup>, which have shown immortalization of MEFs harbouring compound mutations of genes regulating the Rb pathway without obligate loss of p53 pathway function.

The oncogene cooperation assay was used to assess the impact of p16<sup>Ink4a</sup> on cellular transformation (Fig. 3a). In contrast to *Ink4a*/Arf<sup>-/-</sup> cells, p16<sup>Ink4a</sup> deficiency did not enable focus formation by transfection of Ras alone. The combination of Myc and Ras, however, produced a twofold increase in foci in p16<sup>Ink4a</sup>-/- MEFs (Fig. 3a; *P* = 0.037), but this modest increase was far less than that achieved in *Ink4a*/Arf<sup>-/-</sup> cells, even after normalizing for transfection efficiency. Furthermore, as expression of p19<sup>Arf</sup> and p53 was not tested in the transformed foci, our data do not indicate that p16<sup>Ink4a</sup> loss can obviate the need for loss of the p19<sup>Arf</sup>-p53 pathway in this assay. These results indicate that integrity of the p19<sup>Arf</sup>-p53 pathway is the major determinant governing *in vitro* transformability of early-passage MEFs by Myc and Ras, but that a less formidable barrier also exists that can be bypassed by p16<sup>Ink4a</sup> loss.

To assess the tumour-suppressor activity of p16<sup>Ink4a</sup> *in vivo*, a cohort of mice from p16<sup>Ink4a</sup> heterozygous intercrosses was generated and observed for the development of spontaneous malignancy. In animals aged 28–58 weeks (average 44 weeks), 0 out of 26 p16<sup>Ink4a</sup>+/- mice, 4 out of 60 +/- mice, and 10 out of 39 p16<sup>Ink4a</sup>-/- mice had developed tumours (Fig. 3b; *P* = 0.008 for p16<sup>Ink4a</sup>+/- versus -/-). The malignancies in p16<sup>Ink4a</sup>-/- mice were either soft-tissue sarcoma (5), splenic lymphoma (4) or melanoma (1). p19<sup>Arf</sup>-specific and *Ink4a*/Arf-null strains<sup>5,18</sup> seem to exhibit a shorter latency for the development of spontaneous malignancy, although rigorous comparisons in congenic strains are not yet possible. Mice deficient for p16<sup>Ink4a</sup> do seem to be more tumour prone than BALB/c mice, which express a hypomorphic variant<sup>19</sup> of p16<sup>Ink4a</sup> and are not markedly tumour prone compared with other genetic backgrounds. This suggests that the BALB/c p16<sup>Ink4a</sup> allele is not null but hypomorphic as suggested previously<sup>20</sup>, and/or that other unidentified BALB/c loci confer tumour resistance.

Deficiency in p16<sup>Ink4a</sup> has been suggested to render rodents susceptible to certain carcinogens<sup>19,21–24</sup>. In particular, the BALB/c p16<sup>Ink4a</sup> allele seems to confer susceptibility to pristane-induced plasmacytoma and urethane-induced lung adenoma, although these results are confounded by a co-segregating p19<sup>Arf</sup> polymorphism<sup>19,23</sup>. These observations led us to examine whether p16<sup>Ink4a</sup> modulates carcinogen-induced malignancy. Mice from

p16<sup>Ink4a</sup>+/- intercrosses were treated with 7,12-dimethylbenzanthracene (DMBA) as previously described<sup>5</sup>. p16<sup>Ink4a</sup>-/- animals were susceptible to DMBA-induced malignancy, with 50% of the mice developing tumours by 23 weeks compared with >35 weeks for p16<sup>Ink4a</sup>+/- littermate controls (Fig. 3c; *P* = 0.008). A variety of tumours was seen in DMBA-treated animals, including lymphoma, lung adenoma, malignant spindle-cell neoplasm, soft-tissue sarcoma (for example, rhabdomyosarcoma and angiosarcoma), and melanoma (Table 1, Supplementary Information Fig. 7). The spindle-cell neoplasms could represent either poorly differentiated squamous-cell carcinoma reported after DMBA treatment<sup>21,25</sup>, or poorly differentiated sarcoma. Two melanomas, identified by morphology, pigmentation and/or S100 staining, were seen in the p16<sup>Ink4a</sup>-/- colony. Although lung adenomas were noted in nearly all treated mice that died over the age of 14 weeks, these tumours were more clinically aggressive in the p16<sup>Ink4a</sup>-/- mice. That is, p16<sup>Ink4a</sup>-/- mice were more likely to succumb to lung adenoma (because of tachypnea or haemothorax), and only these tumours are included in Table 1. Likewise, skin papillomas were observed with equal frequency in animals of all three genotypes (suggesting comparable carcinogen exposure and metabolism of DMBA), but an increase in malignant spindle-cell neoplasms was seen in the p16<sup>Ink4a</sup>-/- mice compared with +/- controls (Table 1). Therefore, our data suggest that p16<sup>Ink4a</sup> limits progression of pulmonary and skin tumours, and are consistent with the suggestion that p16<sup>Ink4a</sup> loss is associated with the progression from well differentiated squamous-cell carcinoma to poorly differentiated spindle-cell neoplasm as suggested in other models of DMBA-induced skin carcinogenesis<sup>21</sup>.

To assess the role of p16<sup>Ink4a</sup> in another carcinogen model, we treated 23 female littermate mice with urethane as described<sup>23</sup>. Consistent with previous observations<sup>23</sup> in BALB/c hybrid strains, p16<sup>Ink4a</sup> loss conferred about a fourfold increase in number of lung adenomas (Fig. 3d; *P* = 0.02). Also, a trend towards increased nodule size was seen in p16<sup>Ink4a</sup>-/- compared with +/- controls (Fig. 3d), but this difference was not statistically significant (*P* = 0.23). These data, combined with results from other rodent and human studies, suggest that p16<sup>Ink4a</sup> functions to impede malignant progression after exposure to several seemingly diverse carcinogens, including DMBA, urethane, pristane, tobacco and ultraviolet radiation.

Because DMBA-treated p16<sup>Ink4a</sup>+/- animals were more tumour prone than p16<sup>Ink4a</sup>+/- controls, we sought to determine the status of the functional p16<sup>Ink4a</sup> allele in tumours from these mice. Expression of p16<sup>Ink4a</sup> protein was not detected in any tumour from a DMBA-treated animal (2 wild-type and 6 heterozygous mice), whereas p19<sup>Arf</sup> was detected in 4 out of 13 tumours. In all p16<sup>Ink4a</sup>+/- tumours tested, Southern analysis failed to reveal deletions or rearrangements of the wild-type allele (*n* = 22; data not shown), suggesting epigenetic silencing by promoter methylation. We adapted a PCR-based method<sup>26</sup> to detect methylation of CpG sites flanking the translation start site of murine p16<sup>Ink4a</sup>. This assay detected methylation in 7 out of 17 primary tumours as well as in 4 out of 4 cell lines derived from p16<sup>Ink4a</sup>+/- tumours (Fig. 4a). An increase in p16<sup>Ink4a</sup> messenger RNA could also be induced in these cell lines after treatment with 5-aza-2'-deoxycytidine (DAC), a methylation inhibitor (Fig. 4b), demonstrating the correlation of CpG methylation and gene silencing. Interestingly, evidence of promoter methylation was seen in only 2 out of 13 nodules from urethane-treated p16<sup>Ink4a</sup>+/- or +/- mice (Fig. 4a), suggesting that p16<sup>Ink4a</sup> loss may be important in progression rather than initiation of these lesions. As our PCR-based assay only interrogates five CpG sites, however, we cannot exclude the possibility of a failure to detect silencing of the wild-type allele in some cases owing to methylation of other CpG sites.

Our data demonstrate that p16<sup>Ink4a</sup> loss promotes thymic hyperplasia, enhances T-cell proliferative responses and leads to increased

**Table 1 Tumour spectra in DMBA-treated mice**

| Genotype                 | Tumour rate | TL     | SL     | STS    | MSCN   | Lung   | Mel.  | Pap.* |
|--------------------------|-------------|--------|--------|--------|--------|--------|-------|-------|
| p16 <sup>Ink4a</sup> +/- | 6/18        | 3 (17) | 1 (6)  | 1 (6)  | 0      | 1 (6)  | 0     | (33)  |
| p16 <sup>Ink4a</sup> +/- | 20/48       | 6 (13) | 3 (6)  | 5 (10) | 3 (6)  | 3 (6)  | 0     | (29)  |
| p16 <sup>Ink4a</sup> -/- | 23/29       | 6 (21) | 3 (10) | 4 (14) | 4 (14) | 4 (14) | 2 (7) | (24)  |

Percentage of total tumours shown in parentheses. TL, thymic lymphoma; SL, splenic lymphoma; STS, soft-tissue sarcoma; MSCN, malignant spindle-cell neoplasm; Lung, lung adenoma; Mel., melanoma; Pap., squamous papilloma.

\* Only mice that survived more than 16 weeks were considered in the determination of the papilloma frequency.

spontaneous and carcinogen-induced tumours in mice. We also confirm previous results suggesting a more prominent role for p19<sup>Arf</sup> over p16<sup>Ink4a</sup> in mediating passage-induced growth arrest as measured by the 3T9 immortalization assay. The intimate physical arrangement of the coding sequences of these true tumour-suppressor genes helps rationalize the high frequency of deletion of this locus in a variety of human and mouse cancers. These results encourage the question of which protein provides the relevant tumour-suppressor activity of the locus in human cancers. The answer in some tumour types, where homozygous deletion of the locus is common (for example, melanoma and glioblastoma), is almost certainly both p16<sup>Ink4a</sup> and p19<sup>Arf</sup>. As in some of the most common human carcinomas<sup>27</sup>, however, we noted the loss of p16<sup>Ink4a</sup> expression in the setting of promoter methylation. Our data suggest that this epigenetic inactivation occurs frequently and rapidly *in vivo* after carcinogen exposure, and the probability of this event in heterozygous mice is greater than loss of the other allele by deletion. This loss of p16<sup>Ink4a</sup> without structural compromise of the *INK4A/ARF* locus seen in both human and murine tumours indicates that p16<sup>Ink4a</sup> is an important tumour suppressor in certain types of cancer, particularly those associated with carcinogen exposure. □

## Methods

### Targeting construct

The *Ink4a/Arf* locus was cloned from a genomic 129 library. TC1 embryonic stem cells were electroporated and selected as per standard techniques. About 800 clones were screened by Southern blot using a 5' fragment external to the targeting construct (probe A, Fig. 1a) to identify eight recombinants. Blastocyst injections were performed with three different targeted clones, and transmitting chimaeric mice were obtained and bred to FVB/n Cre transgenic mice<sup>6</sup>.

### Colony generation and genotyping

*Cre*<sup>+</sup>p16<sup>Ink4a</sup><sup>+/−</sup> males were crossed to FVB/n females and to littermate *Cre*<sup>+</sup>p16<sup>Ink4a</sup><sup>+/−</sup> mice. p16<sup>Ink4a</sup><sup>+/−</sup> progeny of these matings that were *Cre*<sup>−</sup> were then backcrossed to littermates to yield the bulk of the colony. Most analyses were performed with mice backcrossed two generations to FVB/n. When *Cre*<sup>+</sup> mice were analysed, littermate controls were used that were matched for Cre status and background. Also, all results obtained using *Cre*<sup>+</sup> cells or mice were duplicated in non-transgenic cells and animals. Mice were genotyped by Southern analysis and multiplex PCR (primers and conditions available on request).

### Cellular analysis

Splenocytes were collected from 9-week-old littermates (*n* = 4 per genotype) and mitogenic assays and flow cytometry were performed as previously described<sup>28</sup>. MEFs were generated from embryos at day 13.5 and grown in DMEM with 10% FCS (Hyclone), 50 μM 2-mercaptoethanol, penicillin and streptomycin. For 3T9 analysis, 9 × 10<sup>5</sup> cells were passaged into 6-cm dishes every 3 d. Independent lines (7–9) were assayed per genotype, with two independent cultures per line. In this work a culture was considered 'senescent' if the number of cells decreased (that is, <9 × 10<sup>5</sup>) for two consecutive passages. All lines that had not senesced after 20 passages were immortal, demonstrated by their ability to grow another 40 population doublings and by the absence of senescence-associated β-galactosidase activity. Growth curves were generated by seeding 25,000 cells per well in 12 well plates, each line in triplicate. Plates were fixed on days 0, 2, 4 and 6. Fixed plates were stained with crystal violet and extracted with 10% acetic acid, and relative cell number was measured by absorbance at 595 nm. S-phase re-entry experiments were performed after 72 h of serum starvation.

To measure Ras-induced growth arrest, pBabe or pBabe-activated Hras1 was transfected into 293T cells along with pCL (ref. 29). MEFs were inoculated, selected (4 d in 2 μg ml<sup>−1</sup> puromycin), and assayed as described<sup>12</sup>. For transformation assay, early-passage (passage 4–6) littermate cultures (4–6 lines per genotype) were transfected with pT24–Ras with or without pKO Myc. All transfections were done in duplicate cultures (8 × 10<sup>5</sup> cells) using Lipofectamine Plus in Optimem (Gibco) without serum or antibiotics. Cultures were split 1:3 a day after transfection and foci were counted on day 10. In all experiments, transient transfection efficiency was measured by simultaneous transfection of a replicate plate with green fluorescent protein (GFP). Transfection efficiency did not differ between p16<sup>Ink4a</sup><sup>+/+</sup> and <sup>−/−</sup> cells, but was twofold higher in *Ink4a/Arf*<sup>−/−</sup> cells, and therefore normalized data are reported.

### Western blotting and immunoprecipitation

MEFs (1 × 10<sup>6</sup>) of indicated passage were sonicated in EBC buffer (50 mM Tris pH 8.0, 120 mM NaCl, 0.5% NP40) with protease inhibitors (Complete tabs, Roche) and phosphatase inhibitors (NaF, NaOV). Forty micrograms of lysate were resolved on 8–16% (for p16<sup>Ink4a</sup>, p19<sup>Arf</sup> and p53) or 4–12% (for p53 or Rb) polyacrylamide gels. Membranes were blotted for p16<sup>Ink4a</sup> (M-156, Santa Cruz), p19<sup>Arf</sup> (Novus), p53 (PAB240, Santa Cruz,

or CM5, Novacastra) and Rb (G3245, Pharmingen). Loading was assessed using α-tubulin (Sigma). For p53 induction, lysates were collected 12 h after exposure to ultraviolet radiation (100 mJ m<sup>−2</sup>). Cyclin D–Cdk4 kinase assays were performed as described<sup>30</sup>. Immunoprecipitation of <sup>35</sup>S-labelled cells was done by labelling MEFs for 4 h (750 μCi per 10-cm dish). Cell pellets were lysed in EBC with inhibitors and then 400 μg of lysate were immunoprecipitated with anti-p16<sup>Ink4a</sup> (M-156, Santa Cruz) or anti-Cdk4 (C-22, Santa Cruz) using 1 μg of antibody after preclearing with pre-immune rabbit serum and protein A sepharose. Complexes were resolved by 12% PAGE. For tumours, tissue was dissected from contaminating normal tissue and snap frozen. Lysates were then made by homogenizing ~100 mg of tissue in EBC with inhibitors.

### Tumour cohorts and carcinogen treatment

For the DMBA-treated cohort, 5–7-day-old mice were topically treated with 50 μL of 0.5% DMBA (Sigma). Papillomas were not included in the Kaplan–Meier analysis. Urethane treatment was done as described<sup>23</sup>. Briefly, littermate female mice were treated at 6 weeks old, and analysed 100 d after treatment. Lung nodules were counted and measured in two dimensions. For analysis of methylation status, a p16<sup>Ink4a</sup> specific PCR-based methylation assay was developed after the method of Herman *et al.*<sup>26</sup>. After modification with bisulphite, tumour DNA was then amplified with primers specific for unmethylated (U3 = GTGATTGGGTGGGTATTGAATTTTGTG, U4 = CACACATCATACACACAA CCCTAAACCA), methylated (M3 = CGATTGGGCGGGTATTGAATTTTCGC, M4 = CACGTCATACACACGACCCTAAACCG) or unmethylated sequences (W1 = ACTGAATTC TCCGCGAGGAAAGCG, W2 = GCACACGGCCCTGGGCGCCG) flanking the p16<sup>Ink4a</sup> translation start site. All modified samples did not amplify with primers specific for the unmethylated sequences (data not shown). Cell lines were derived from primary tumours by culturing in RPMI with 10% FCS, 50 μM 2-mercaptoethanol, penicillin and streptomycin.

Received 12 June; accepted 10 July 2001.

- Ruas, M. & Peters, G. The p16<sup>Ink4a</sup>/CDKN2A tumor suppressor and its relatives. *Biochim. Biophys. Acta* **1378**, F115–F177 (1998).
- Sharpless, N. E. & DePinho, R. A. The *INK4A/ARF* locus and its two gene products. *Curr. Opin. Genet. Dev.* **9**, 22–30 (1999).
- Quelle, D. E., Zindy, F., Ashmun, R. A. & Sherr, C. J. Alternative reading frames of the *INK4a* tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* **83**, 993–1000 (1995).
- Kamijo, T. *et al.* Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product p19<sup>Arf</sup>. *Cell* **91**, 649–659 (1997).
- Serrano, M. *et al.* Role of the *INK4a* locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
- Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl Acad. Sci. USA* **93**, 5860–5865 (1996).
- Pham, C. T., MacIvor, D. M., Hug, B. A., Heusel, J. W. & Ley, T. J. Long-range disruption of gene expression by a selectable marker cassette. *Proc. Natl Acad. Sci. USA* **93**, 13090–13095 (1996).
- Zindy, F., Quelle, D. E., Roussel, M. F. & Sherr, C. J. Expression of the p16<sup>Ink4a</sup> tumor suppressor versus other *INK4* family members during mouse development and aging. *Oncogene* **15**, 203–211 (1997).
- Alcorta, D. A. *et al.* Involvement of the cyclin-dependent kinase inhibitor p16 (*INK4a*) in replicative senescence of normal human fibroblasts. *Proc. Natl Acad. Sci. USA* **93**, 13742–13747 (1996).
- Rogan, E. M. *et al.* Alterations in p53 and p16<sup>Ink4a</sup> expression and telomere length during spontaneous immortalization of Li-Fraumeni syndrome fibroblasts. *Mol. Cell. Biol.* **15**, 4745–4753 (1995).
- Reznikoff, C. A. *et al.* Elevated p16 at senescence and loss of p16 at immortalization in human papillomavirus 16 E6, but not E7, transformed human uroepithelial cells. *Cancer Res.* **56**, 2886–2890 (1996).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16<sup>Ink4a</sup>. *Cell* **88**, 593–602 (1997).
- Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
- Carnero, A., Hudson, J. D., Price, C. M. & Beach, D. H. p16<sup>Ink4a</sup> and p19<sup>Arf</sup> act in overlapping pathways in cellular immortalization. *Nature Cell Biol.* **2**, 148–155 (2000).
- Dannenberg, J. H., van Rossum, A., Schuijff, L. & te Riele, H. Ablation of the *retinoblastoma* gene family deregulates G(1) control causing immortalization and increased cell turnover under growth-restricting conditions. *Genes Dev.* **14**, 3051–3064 (2000).
- Sage, J. *et al.* Targeted disruption of the three Rb-related genes leads to loss of G(1) control and immortalization. *Genes Dev.* **14**, 3037–3050 (2000).
- Groth, A., Weber, J. D., Willumsen, B. M., Sherr, C. J. & Roussel, M. F. Oncogenic Ras induces p19<sup>Arf</sup> and growth arrest in mouse embryo fibroblasts lacking p21<sup>Cip1</sup> and p27<sup>Kip1</sup> without activating cyclin D-dependent kinases. *J. Biol. Chem.* **275**, 27473–27480 (2000).
- Kamijo, T., Bodner, S., van de Kamp, E., Randle, D. H. & Sherr, C. J. Tumor spectrum in *ARF*-deficient mice. *Cancer Res.* **59**, 2217–2222 (1999).
- Zhang, S., Ramsay, E. S. & Mock, B. A. Cdkn2a, the cyclin-dependent kinase inhibitor encoding p16<sup>Ink4a</sup> and p19<sup>Arf</sup>, is a candidate for the plasmacytoma susceptibility locus, Pctr1. *Proc. Natl Acad. Sci. USA* **95**, 2429–2434 (1998).
- Zhang, S. L. *et al.* Efficiency alleles of the Pctr1 modifier locus for plasmacytoma susceptibility. *Mol. Cell. Biol.* **21**, 310–318 (2001).
- Linaropoulos, S. *et al.* Deletion and altered regulation of p16<sup>Ink4a</sup> and p15<sup>Ink4b</sup> in undifferentiated mouse skin tumors. *Cancer Res.* **55**, 5168–5172 (1995).
- Swafford, D. S. *et al.* Frequent aberrant methylation of p16<sup>Ink4a</sup> in primary rat lung tumors. *Mol. Cell. Biol.* **17**, 1366–1374 (1997).
- Herzog, C. R., Noh, S., Lantry, L. E., Guan, K. L. & You, M. Cdkn2a encodes functional variation of p16<sup>Ink4a</sup> but not p19<sup>Arf</sup>, which confers selection in mouse lung tumorigenesis. *Mol. Carcinog.* **25**, 92–98 (1999).
- Patel, A. C. *et al.* Hypermethylation of the p16 (*Ink4a*) promoter in B6C3F1 mouse primary lung adenocarcinomas and mouse lung cell lines. *Carcinogenesis* **21**, 1691–1700 (2000).



25. Buchmann, A., Ruggeri, B., Klein-Szanto, A. J. & Balmain, A. Progression of squamous carcinoma cells to spindle carcinomas of mouse skin is associated with an imbalance of H-ras alleles on chromosome 7. *Cancer Res.* **51**, 4097–4101 (1991).
26. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA* **93**, 9821–9826 (1996).
27. Herman, J. G. *et al.* Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* **55**, 4525–4530 (1995).
28. Carrasco, D. *et al.* Multiple hemopoietic defects and lymphoid hyperplasia in mice lacking the transcriptional activation domain of the c-Rel protein. *J. Exp. Med.* **187**, 973–984 (1998).
29. Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
30. Matsushime, H. *et al.* D-type cyclin-dependent kinase activity in mammalian cells. *Mol. Cell. Biol.* **14**, 2066–2076 (1994).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

## Acknowledgements

We thank E. Farazi, S. Chan, M. Sinha, S. Alson and the Dana Farber Cancer Institute Mouse Core for experimental advice and assistance. We also thank H. Westphal, T. Devereux, M. Oren, M. Bosenberg, D. Silver and A. Buchmann for advice or reagents, and L. Chin, J. DeCaprio, W. Kaelin and P. Sicinski for comments on the manuscript. N.E.S. is an Howard Hughes Medical Institute Physician Postdoctoral Fellow; N.B. is supported by the American Cancer Society John Peter Hoffman Award; K.H.L. is supported by the National Nuclear R and D Program and Human Genome Program from the Korean Ministry of Science and Technology; D.H.C. is supported by a Physician Scientist Fellowship of the Damon Runyon–Walter Winchell Foundation; and R.A.D. is an American Cancer Society Professor and recipient of the Steven and Michele Kirsch Foundation Investigator Award. This work was supported by grants from the National Institutes of Health, the National Cancer Institute and the American Cancer Society.

Correspondence and requests for materials should be addressed to R.A.D. (e-mail: [ron\\_depinho@dfci.harvard.edu](mailto:ron_depinho@dfci.harvard.edu)).

## Contacts

**Publisher:** Fabien Savenay  
**Editor:** Paul Smaglik  
**Sales Director:** Ben Crowe

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [classified@nature.com](mailto:classified@nature.com)

### Group European Manager:

Leonie Wells (4954)

### European Manager:

Nevin Bayoumi (4978)

### UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),  
Andy Douglas (4975)

**Holland/ Italy:** Nevin Bayoumi (4978)

**Scandinavia:** Silje Opstrup (4994)

**Spain/ Portugal:** Leonie Wells (4954)

### Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [classified@nature.com](mailto:classified@nature.com)

### International

#### Advertising Coordinator:

Laura Pearson (4977)

#### Naturejobs web development:

Tom Hancock

#### Naturejobs online production:

Ben Lund

### European Satellite Offices

#### France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: [c.nioxchateau@nature.com](mailto:c.nioxchateau@nature.com)

#### Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: [p.phelan@nature.com](mailto:p.phelan@nature.com)

[k.turner@nature.com](mailto:k.turner@nature.com)

### US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: [classified@natureny.com](mailto:classified@natureny.com)

### US Sales Director:

Ben Crowe

### US Sales Manager:

Peyton Mason

### US Advertising Coordinator:

Corissa Salzman

### Japan Head Office, Tokyo

MG Ichigaya Building (6F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: [k.cowan@naturejpn.com](mailto:k.cowan@naturejpn.com)

### Japan Manager:

Kate Cowan



## Presenting Naturejobs.com 2

When websites are being redesigned and retooled, they are frequently flagged with signs saying 'under construction'. Such construction usually means temporary inconvenience, with the implicit promise that, once all the potholes are patched, use will become faster and easier.

For the past several months, naturejobs.com (<http://www.nature.com/naturejobs>) has undergone such roadwork. During that time, efforts to improve the site's features, such as job searching and résumé posting, have meant that users have had to endure detours.

Today, the barricades have been lifted, allowing users new routes to expanded services. These services include more refined job searches, easier ways to post electronic résumés, and e-mail alerts about suitable jobs.

Construction has extended to the way *Naturejobs* organizes its editorial content. The main page still contains the most recent articles that examine how science and policy are changing employment dynamics in different disciplines around the world.

But to make the collection of articles more relevant to different interests, we've separated them into life sciences and physics. In doing so, we've supplemented them with other resources from within *Nature*. The life-sciences 'channel' now also includes material from *Nature Biotechnology*, and the physics 'channel' is closely linked with *Nature's* physics portal. In addition, both of those sites contain links to outside resources that we think will be useful.

We hope you agree that the better roads to our improved services have been worth the wait.

### Paul Smaglik

*Naturejobs* editor



## Contents

### CAREERS AND RECRUITMENT

Structural biology faces industrial revolution p4

Japan invests in structural genomics p7

### WWW.NATURE.COM/ NATUREJOBS

Career centre  
Information on the scientific job market

#### FOCUS

#### SPOTLIGHT

#### RECRUITMENT

#### SCIENTIFIC ANNOUNCEMENTS

#### SCIENTIFIC EVENTS

# Building protein pipelines

The industrial revolution that is reshaping structural biology is opening doors of opportunity for computer scientists and protein chemists alike, says Paul Smaglik.



Ian Wilson (left) believes protein chemists are in big demand.

**A**utomation promises the same dramatic results in protein science as the introduction of the assembly line produced for the textile industry. By speeding the way in which the three-dimensional shapes of proteins are discerned, the world's protein-research centres will become high-speed discovery factories (see *Nature* **408**, 130–132; 2000).

The movement to speed every aspect of structural biology — selecting a target protein, expressing it, determining its shape and annotating it — will require the skills of chemists, engineers and computer scientists. If successful, the move towards automation will reduce the need for scientific labour at the beginning of the line, but increase work for biologists at the end.

This holds true for structural biologists working in academia and industry. And the players in all sectors are already carving out their niches. Broadly speaking, public projects are leading in technology development, pharmaceutical companies are using available structural information to optimize leads and improve existing compounds, and biotech companies are searching for even more specialized roles.

## ACADEMIC LEADERS

Large public projects such as the US structural-genomics initiatives — on which the National Institute of General Medical Science is spending \$30 million a year over five years — may well provide the engine that drives the change. The US project's long-term goal is to determine the structures of 10,000 different proteins in 10 years (see *Nature* **407**, 549; 2000). But the first five years of the project will be spent developing technology to enable the project's centres to achieve that aim in the remaining five.

Accordingly, Ian Wilson, director of the Joint Center for Structural Genomics at the Scripps Research Institute in La Jolla, California, has so far not

## Solving a career problem

**José Cosme discovered a good way to get a permanent position in structural biology — help to solve the structure of a difficult protein that plays a key role in drug metabolism.**

**When Cosme was in his last year of a postdoc at the Scripps Research Institute in La Jolla, California, he was part of a team that cracked the first structure of a mammalian cytochrome P450. It was a hard task because the protein is a membrane protein with lots of hydrophobic regions, which makes it resistant to crystallization.**

**Slight variations in this protein's structure can have profound**

**consequences on drug uptake.**

**Within a month of Cosme's team publishing a paper describing the rabbit version of the protein (*Mol. Cell* **5**, 121–131; 2000), the co-founder of structure-based drug-design company Astex flew from England to recruit him. He asked Cosme to join the company and help to solve the human version of the protein.**

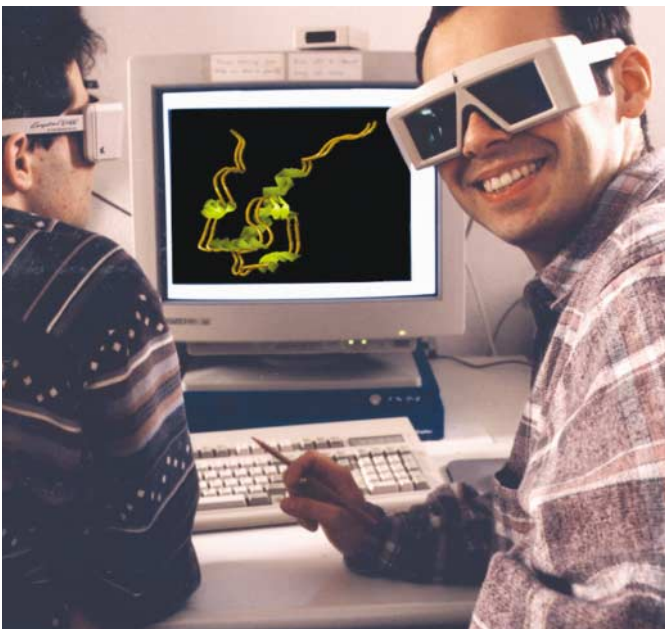
**The company has since upped its commitment to the protein, by recruiting the paper's first author, Pamela Williams. Cosme did much of the protein chemistry work, and Williams crystallized it.**

**P.S.**

been hiring many biologists. "We've hired project managers like you'd have in mega physics projects," Wilson says. He has put in place management for bioinformatics, 'crystallography' — which includes the stages from protein expression to crystallization — and structural determination.

Wilson is also keen to consolidate and expand the centre's team of engineers and technicians with collaborators at the Genomics Institute of the Novartis Research Foundation in San Diego and the





PROTEIN STRUCTURE FACTORY

biotech company Syrrx, also in San Diego. “We’re trying to assemble an automated pipeline to do high-throughput structural determination with essentially as little manual intervention as possible,” Wilson says. Scientists with specialized skills who can anticipate and prevent blocks in the pipeline will also be in high demand.

The biggest technological bottleneck will be producing a high enough percentage of soluble

protein. At the moment, even under ideal conditions, only about 20% of proteins are soluble. And the success ratio goes down considerably as the size of the target protein goes up — especially if the targets include membrane proteins. This is key because, even though they account for about a quarter of all known proteins, membrane proteins make up close to half of the targets that drugs companies are interested in. This means that protein chemists will be in big demand, says Wilson.

#### PHARMA CAUTIOUS

Although pharmaceutical companies are eager to exploit structures for drug design, they do not want to spend a lot of money building up infrastructure in the way that the public projects are doing. And they are more cautious about the applied promise of the basic science. Years ago, the conventional wisdom was that structure could be used to develop drugs from the ground up. That has proven to be the exception rather than the rule.

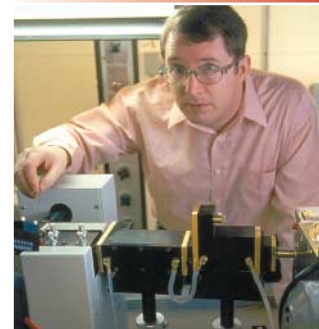
Big companies are now focusing on using existing structures to refine drugs already in the pipeline, says Peter Kim, executive vice-president of research and development at Merck in New Jersey. But the company has slowly increased its activities in structural biology, and this is likely to continue as it brings in more X-ray crystallographers, molecular-modelling experts and protein-expression specialists, he says.

The ability to design drugs from structure will require more basic science — especially about the complex chemical interactions that occur between two molecules. But calculating the energy of such interactions and the binding affinity between molecules is not the kind of work supported by



MERCK

Merck's Peter Kim expects the company to engage in more structural biology.



On the up: AstraZeneca has expanded its protein chemistry effort over the past two years.

ASTRAZENECA

## Automation's crystal ball

It seems ironic that those behind the projects designed to produce an abundance of new protein structures are hiring relatively few stalwarts of structural biology. Indeed, the largest US structural-genomics initiative is not funding any postdocs, because the project's focus will provide little useful training for fledgling structural biologists.

‘Discovery science’ has come to structural biology — and the employment and funding of scientists in the field promises to be greatly altered. Publicly funded initiatives will initially each produce hundreds of protein structures a year. As they scale up, the number could grow to thousands — eventually providing examples of most of the 30,000 or so families of protein folds that

occur in nature.

So will this burst of productivity put structural biologists out of work? On the contrary, say leaders in academia and industry. Drugs companies will still need their own structural experts because many of the structures generated by public efforts may be scientifically interesting, but not immediately applicable.

Karl-Heinz Altmann, a chemist at Novartis, says that the high-throughput approach favoured by most public projects may be of little initial relevance to large drugs companies. “We’re normally interested in a specific protein,” Altmann says. “It’s not very helpful to us if we can crystallize protein A when we are interested in protein B.”

Some public projects select target proteins using different criteria from industry. They want to cover a wide variety of protein shapes, whereas industry is more interested in proteins that are known to play a role in key pathways. And the commercial sector is also interested in large molecules that are most resistant to a high-throughput approach, such as membrane proteins.

Also, having a structure in the databank from one protein family does not ensure that it will be relevant to developing a drug based on the structure of its relative.

Nigel Darby, vice-president of enabling science and technology (chemistry) at AstraZeneca in Mölndal, Sweden, adds that drugs

companies are often more interested in the structures of proteins with different compounds bound to them than in the structures of the proteins alone.

Ian Wilson, director of the Joint Center for Structural Genomics at the Scripps Research Institute in La Jolla, California, agrees that the approaches taken by public and private initiatives are different. But he believes that the sheer quantity of structures will bridge any gaps. “What is going to get transferred to the community is an enormous number of protein structures for which there is no biology,” he says. And that will create a lot of exciting opportunities for molecular biologists, in both academia and industry. **P.S.**

pharmaceuticals — it is far too basic, Kim says. “I would love to see more funding for it.”

Kim, like Wilson, sees the more immediate need for applied biochemists. The people most in demand, he says, are those who are adept at protein expression. “What one has to have in every drug-discovery programme is properly folded, pure, well-characterized proteins,” Kim says.

Protein wizards are in short supply, agrees Nigel Darby, vice-president of enabling science and technology (chemistry) at AstraZeneca in Mölndal, Sweden. He says that a good protein chemist’s abilities should extend beyond just producing proteins. “Protein engineering to modify the protein is often required to increase the chances of crystallization or to generate good NMR spectra,” he says.

AstraZeneca has rapidly built up its structural chemistry team — it has recruited 35 people to its structural biology/protein chemistry effort in the past two years. Recruits have included X-ray crystallographers, NMR experts and molecular biologists who specialize in structure. Perhaps the biggest growth area is in computational chemistry — people who can use both structural and chemical data and apply them to drug design. These sorts of individuals belong to a growing pool of hybrids who are attractive to industry. “People need multiple skill sets — or at least an ability and willingness to acquire new skills,” Darby says.

Stephen Burley: from ‘computer nerd’ to structural biologist.



## An odd road to structural biology

**Stephen Burley of Rockefeller University, New York, and leader of one of the US structural-genomics initiative’s seven centres, could not have picked a more twisted career path to his current position.**

**As an undergraduate “I was a computer nerd”, he says. Although interested in the quantitative sciences, biology caught his eye, so he did a doctorate at**

**Oxford in molecular biophysics.**

**Needing more training in biology, he went to Harvard Medical School. “I loved patient care,” he says. But he also loved research. He soon realized that he had to decide between balancing research and care or doing research full time. He postponed the decision by doing a research residency at Harvard-affiliated Brigham and Women’s Hospital. The stint allowed him to split the difference of his interests — six months in the lab, six months on the wards. “I discovered I couldn’t do both,” Burley says.**

**Solving the structure of the TATA box protein (see *Nature* 360, 40–46; 1992) put him on the road to full-time structural biologist. In retrospect, his path to leadership in structural genomics seems obvious.**

P.S.

ROCKEFELLER UNIV.

### BIOTECH NICHE

Chemical skills are also in demand in biotech companies that are basing some of their business plans on structural biology, says Raymond Salemme, founder, president and chief scientific officer of 3-Dimensional Pharmaceuticals (3DP) in Exton, Pennsylvania. Salemme believes that mixing combinatorial chemistry and structural biology will produce better drugs.

But his company’s biggest need is for biophysicists. The drug-screening strategy at 3DP involves measuring the binding of biological cofactors to genomic targets. Once the affinity has been determined, biologists can unravel a protein’s function based on what binds to it. That information, in turn, can be applied to refining the protein-chemistry techniques necessary to crystallize the protein.

Structural GenomiX in San Diego, on the other hand, is setting itself up as a mini version of the US structural-genomics initiative. It is building its own synchrotron beamline, used to help elucidate a protein’s shape, at Argonne National Laboratory. Many labs and companies must wait in line for beam time,

so having its own will help the company skip the queue. At the moment, Structural GenomiX is seeking technical people to assist in automation. Once the beamline is complete, the company will start recruiting structural biologists.

British company Astex is taking, if anything, the opposite approach. It is choosing target proteins that it thinks will be attractive to industry and focusing on those, says Harren Jhoti, chief scientific officer and co-founder of the Cambridge-based company. Its recruitment strategy matches its scientific and business schemes.

And once targets have been selected, the company will turn to third-party technology providers, rather than try to do all the automation itself. Although eager to benefit from the structural biology industrial revolution, Astex, like many larger companies, is happy to let someone else build the factory. ■

Paul Smaglik is editor of *Naturejobs*.

### Web links

NIGMS structural-genomics initiative

▶ <http://www.nigms.nih.gov/funding/psi.html>

Joint Center for Structural Genomics

▶ <http://www.jcsg.org/home.html>

Human Proteome Organisation ▶ <http://www.hupo.org>

# Japan in the post-genomics age

Over the past two years, the Japanese government has increased funding for research in structural genomics. Although companies remain uncertain as to how this will benefit them, there will be more job opportunities in protein engineering and structural biology, says Robert Triendl.



Shape of things to come: structural genomics is a major focus at RIKEN's Genomic Sciences Center.

consortium to build and operate a beamline dedicated to protein-structure determination at Japan's SPring-8 synchrotron radiation facility were amazed earlier this year when 21 companies joined. Toshikazu Miyagishima, manager of the healthcare information division at Tanabe Seiyaku and chair of the Japan Pharmaceutical Manufacturers Association's R&D committee, says he expected fewer than a dozen.

The unexpectedly large number of participants illustrates the rapid surge of interest in structural genomics and protein-structure determination in the Japanese pharmaceutical industry, says Miyagishima. The new beamline will provide companies with access to some of the most advanced instrumentation for structural genomics presently available.

## TARGET IDENTIFICATION

Most companies in Japan say they plan to use high-throughput structure-determination methods to investigate favourite drug targets, and only a few companies say they will attempt to investigate new targets, which is a much more demanding task. This seems to confirm the observation, made by many industry analysts, that only a handful of pharmaceutical companies in Japan can support full-scale drug-development programmes based on structural genomics.

Structural genomics tends to be seen as just another tool to add to most companies' existing repertoire of research techniques, rather than a fundamentally new approach. In particular, smaller pharmaceutical companies will use structural-genomics tools initially to narrow down the number of leads, to refine lead compounds already under development, or to re-evaluate compounds that had dropped out at a late stage of the drug-development process. But several pharmaceutical companies are cooperating with RIKEN, the Japanese Institute of Physical and Chemical Research, in an international

RIKEN

**T**here is little doubt that Japan has missed the genomics revolution, but there seems to be a widespread belief among Japanese scientists and industrialists that the country is in a much better position to capitalize on structural genomics. The Japanese government has made large increases in funding for structural genomics explicitly to help the Japanese pharmaceutical industry enter the post-genomics age.

But industry spokespeople say that it is not yet clear to many companies how they will benefit from the new efforts. Future success in applying structural genomics and high-throughput structure-determination technologies to drug development will depend on cooperation between the public sector, specialized biotechnology or software companies and the pharmaceutical industry.

Research on protein-structure determination is rapidly emerging as a major priority in post-genomics research in Japan. The organizers of an industry



effort to determine 3,000 new protein structures.

Although industry remains sceptical about large-scale efforts to determine thousands of new protein structures, structural genomics has rapidly emerged over the past two years as a major recipient of funding in the life sciences and biotechnology in Japan. Structural genomics is a major research focus at the RIKEN Genomic Sciences Center (GSC), Japan's foremost genomics research centre, and the Japan Biological Information Research Center (JBIRC), a new centre for structural genomics funded by the Ministry of Economy, Trade and Industry (METI), which opened its doors in April.

The structural-genomics initiative at RIKEN consists of two large projects — the structurome project and the protein-fold project. Both efforts are funded at around US\$5 million to \$7 million per year over several years, with additional funding provided for instruments and facilities. The structurome project, located at RIKEN's Nishi-Harima Research Center, is an effort to determine the structure of all proteins of the extremophile bacterium *Thermus thermophilus*. Although the structurome project will mainly use X-ray crystallography for structure determination, the protein-fold project at the GSC has been intimately linked since its inception to GSC's large new nuclear magnetic resonance facility. Research activities under the protein-fold project will focus on structure determination of mouse and plant proteins, using complementary DNA libraries developed by scientists at the GSC.

Online success: Japan's SPring-8 synchrotron facility has attracted unprecedented interest.

#### INFORMATION INITIATIVES

Research at the JBIRC will focus mostly on human transmembrane proteins, which are not typically a focus of structural-genomics projects — membrane proteins are difficult to obtain in sufficient quantities and do not crystallize easily. At its completion, the JBIRC expects to host some 100 full-time researchers in three divisions, specializing in structural genomics, functional genomics and bioinformatics. An overall budget of ¥3.5 billion (US\$32 million) per year for a seven-year period was approved by METI. But although the JBIRC is part of METI's National Institute for Advanced Industrial Science and Technology, most of the funding comes from the New Energy and Industrial Technology Development Organization, a METI funding body. As a result, only some 20 of the positions available at the JBIRC confer government official status, and for the remaining positions only temporary contracts will be available.

To aid interactions with industry, management of the JBIRC has been entrusted to the Japan Biological Informatics Consortium (JBiC), an industry organization with some 83 corporate members from the life sciences and information-technology industries. Although the JBIRC plans to hire new staff over the next few months, some of the positions available will also be filled by scientists from JBiC member companies or universities. But filling the new positions is likely to take more time than initially expected — for the centre's director, Yoshimasa Kyogoku, this is an unusual situation. "At the university we used to have a large number of students. Here we have plenty of funding, but as yet only limited staff," he says.

One of the goals behind setting up the JBiC was to jump-start an indigenous bioinformatics industry. But even today the Japanese market for computational tools or databases in structural genomics remains dominated by US and European companies, although several start-up companies in Japan — including PharmaDesign and Protein Express — now offer software products and services for structure prediction, structure databases or rational drug design. Large information-technology companies, such as NEC, have also announced their intention to accelerate the development of high-performance systems and software tools for knowledge-based protein-structure prediction or molecular-dynamics simulations for drug-development applications such as the 'virtual screening' of compounds.

Larger pharmaceutical companies such as Yamanouchi and Eisai say that they plan to increase investment in genomics and post-genomics research significantly over the next few years — and research on protein structure is rapidly emerging as a major focus at many companies. Other companies are also gearing up activities in protein engineering, structural genomics and proteomics.

Even outsiders such as the consumer electronics company Sanyo and the automotive company Honda



have started to hire scientists with a background in biotechnology, genomics or protein engineering.

Specialized recruiters say employment opportunities for scientists with a background in protein expression, structure determination, protein bioinformatics or rational drug design are likely to increase over the next few years, but that this may not always mean permanent employment. In trying to catch up in the genomics field, many pharmaceutical companies in Japan have started to hire more mid-career scientists or have contracted specialized staff through temporary employment or staff-outsourcing agencies, such as Hitec. And, especially in Japan, temporary employment is still considered second-rate. ■

Robert Triendl is a freelance writer based in Tokyo.

#### Web links

SPring-8 ♦ <http://www.spring8.or.jp>

RIKEN ♦ <http://www.riken.go.jp>

Japan Biological Informatics Consortium ♦ <http://www.jbic.or.jp>

RIKEN